A Corpus Study of Differences in the Spoken English and Spoken Mandarin Chinese of Taiwanese EFL Students in Liverpool and Taiwan

Thesis submitted in accordance with
the requirements of the University of Liverpool
for the degree of Doctor of Philosophy
by
Wan-Ting Chen

December 2017

Abstract

Wan-Ting Chen

A Corpus Study of Differences in the Spoken English and Spoken Mandarin Chinese of Taiwanese EFL Students in Liverpool and Taiwan

This thesis is concerned with whether the Taiwanese EFL learners with different English environment/education backgrounds have noticeably different language use from each other in spoken English and spoken Mandarin. This thesis aims to investigate the differences in language use of two Taiwanese EFL learner communities in the UK (Liverpool) and Taiwan (Taipei). The Taiwanese EFL learner community in Taiwan is further divided into two groups because of their different educational backgrounds (English-relevant subjects and non-English-relevant subjects.) Our investigation looks at the use by the three groups of English pronouns *I*, *you* and *it*, Mandarin pronouns 我 (wǒ) [I], 你 (nǐ) [you] and 它 (tā) [it]. Their use of English verbs *is* and *have* and the Mandarin verbs 是 (shì) [to be] and 有 (yǒu) [to have/exist] are also investigated. The investigation focuses on whether there are significant differences in the use of the highly frequent L1 and R1 collocates of these words and the patterns in which these collocations are used. We investigate whether these differences are potentially explained by the influence of the different kinds of input from the environment and educational training to which these EFL learners are exposed. Our findings show that significant differences in the use of spoken English between the three groups exist. Differences in the nature of the English input to which the three groups have been exposed appear to be a possible explanation of differences in these participants' English use. However, we also find there are significant differences in the use of spoken Mandarin between the groups of participants with different English input backgrounds. What we identify as differences suggests that a cross-language influence on people's first and second/foreign language use may exist.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisors, Prof Michael Hoey and Dr Hitomi Masuhara for valuable comments and suggestions. They have been helpful and enlightening at every stage of my PhD study. This thesis could not have been accomplished without their kind and encouraging guidance.

Next, I wish to thank my MA supervisor Dr Michael Pace-Sigge. I would never have thought of undertaking PhD study if he had not encouraged me in his supervision of my MA dissertation. I also wish to thank my dear teacher, Dr Su-Hsun Tsai, for her assistance. Without her kind assistance and care, I would not have studied at the University of Liverpool and completed my study.

Finally, I am grateful to my family and friends in Taiwan, whose encouragement and support gave me the strength to complete my study abroad.

Table of Contents

Abstract i

Acknowledgements ii

Table of Contents iii

Index of Tables vi

Chapter 1 Introduction 1

1.1 The motivation behind this research 1

1.2 The potential value of this research 7

1.3 The structure of this thesis 8

Chapter 2 Theoretical Background 9

2.1 A brief review of SLA theories 9

2.1.1 Input Hypothesis, Output Hypothesis and Interaction Hypothesis 13

2.2 Language learning, language use and minds 21

2.3 The hypotheses of this study 50

Chapter 3 Methodology 52

3.1 Overview 52

3.2 Building a Taiwanese EFL learner corpus 59

3.2.1 Three groups: LIVT, TESOL and NTESOL 62

3.3 Method of compilation 70

3.3.1 Data Transcription 71

3.3.2 Reference English Spoken Corpora 72

3.3.3 Reference Mandarin Spoken Corpus 74

3.4 Method of Analysis 74

3.4.1 Words for Analysis 74

3.4.2 Procedure of Analysis 76

Chapter 4: Use of I, YOU and IT by the three Taiwanese groups 78

4.1 Introduction 78

4.2 The Use of *I* 83

4.2.1 L1 Collocates of *I* 85

4.2.2 R1 Collocates of *I* 94

4.2.3 Summary of the Use of *I* 106

4.3 The Use of *you* 109

4.3.1 L1 Collocates of *you* 110

4.3.2 R1 Collocates of *you* 124

4.3.3 Summary of the use of *you* 129

4.4 The Use of *it* 132

4.4.1 L1 Collocates of *it* 133

4.4.2 R1 Collocates of *it* 140
4.4.3 Summary of the Use of *it* 144
4.5 Conclusion 146
Chapter 5 Use of 我(wǒ) [I], 你(nǐ) [you] and 它(tā) [it] by the three Taiwanese groups 148
5.1 Introduction 148
5.2 The Use of 我(wǒ) [I] 149
5.2.1 L1 Collocates of 我(wǒ) [I] 149
5.2.2 R1 Collocates of 我(wǒ) [I] 157
5.2.3 Summary of the Use of 我(wǒ) [I] 167
5.3 The Use of 你(nǐ) [you] 169
5.3.1 L1 Collocates of 你(nǐ) [you] 169
5.3.2 R1 Collocates of 你(nǐ) [you] 183
5.3.3 Summary of the Use of 你(nǐ) [you] 192
5.4 The Use of 它(tā) [it] 195
5.4.1 L1 Collocates of 它(tā) [it] 195
5.4.2 R1 Collocates of 它(tā) [it] 203
5.4.3 Summary of the Use of 它(tā) [it] 208
5.5 Conclusion 212
Chapter 6 Use of IS and 是(shì) [to be] by the three Taiwanese groups 216
6.1 Introduction 216
6.2 The Use of *is* 218
6.2.1 L1 Collocates of *is* 220
6.2.2 R1 Collocates of *is* 232
6.2.3 Summary of the use of *is* 245
6.3 The Use of 是(shì) [to be] 248
6.3.1 L1 Collocates of 是(shì) [be] 250
6.3.2 R1 Collocates of 是(shì) [be] 264
6.3.3 Summary of the Use of 是(shì) [be] 278
6.4 Conclusion 281
Chapter 7 Use of HAVE and 有(yǒu) [to have, to exist] by the three Taiwanese groups 287
7.1 Introduction 287
7.2 The Use of *have* 291
7.2.1 L1 Collocates of *have* 291
7.2.2 R1 Collocates of *have* 298
7.2.3 Summary of the Use of *have* 303
7.3 The Use of 有(yǒu) [to have, to exist] 305
7.3.1 L1 Collocates of 有(yǒu) [to have, to exist] 305
7.3.2 R1 Collocates of 有(yǒu) [to have, to exist] 314

7.3.3 Summary of the Use of 有(yǒu) [to have, to exist] 325
7.4 Conclusion 327
Chapter 8 Conclusions 330
8.1 Conclusions of the current study 330
8.2 Limitations of this research and Suggestions for the future research 335
Appendices 339
Appendix 1. Top 30 frequent wordlists 339
Appendix 2. Sample transcriptions and transcription conventions for corpus data 343
Bibliography 351

Index of Tables

Chapter 3

Table 3.1 The EFL learning situations: English as a subject/medium 54

Table 3.2 The EFL learning situations: English input in non-English-relevant subjects 55

Table 3.3 The EFL learning situations: English input in English-relevant subjects 56

Table 3.4 The EFL learning situations: English input in different groups 58

Table 3.5 English Interview Questions 60

Table 3.6 Mandarin Chinese Interview Questions 61

Table 3.7 Background information of the TESOL participants 63

Table 3.8 Background information of the NTESOL participants 66

Table 3.9 Background information of the LIVT participants 67

Table 3.10 Hours of English use after school (LIVT participants) 68

Table 3.11 Extracurricular English Input (LIVT participants) 69

Table 3.12 The sizes of the LIVT, TESOL and NTESOL English corpora 72

Table 3.13 The sizes of the LIVT, TESOL and NTESOL Mandarin Chinese corpora 72

Table 3.14 The size of the BASEah corpus 73

Table 3.15 The size of the COCAsp corpus 73

Table 3.16 The size of the Sinica (spoken) corpus 74

Chapter 4

Table 4.1 Frequencies of *I* and 我(wǒ) in LIVT, TESOL and NTESOL 79

Table 4.2 Frequencies of *you* and 你(nǐ) in LIVT, TESOL and NTESOL 79

Table 4.3 Frequencies of *it/it's* and 它(tā) in LIVT, TESOL and NTESOL 80

Table 4.4 Log-Likelihood Ratios: *I* and 我(wǒ) in LIVT, TESOL and NTESOL 83

Table 4.5 L1 collocates of *I* in LIVT, TESOL and NTESOL 85

Table 4.6 Log-likelihood Ratios: The L1 collocates of *I* in LIVT, TESOL and NTESOL 86

Table 4.7 Frequencies and Log-likelihood Ratios: *because* in LIVT, TESOL and NTESOL 86

Table 4.8 Use of *yeah* and *yes* in LIVT, TESOL and NTESOL (1) 89

Table 4.9 Use of *yeah* and *yes* in LIVT, TESOL and NTESOL (2) 89

Table 4.10 Log-likelihood Ratios: *and I* and *and* 91

Table 4.11 Use of *and* and *and I* in LIVT, TESOL and NTESOL 91

Table 4.12 Log-likelihood Ratios: *yeah*/*yes* and *I* and *yeah/yes and* 91

Table 4.13 Use of *and* and *and I* in BASEah and COCAsp 92

Table 4.14 R1 collocates of *I* in LIVT, TESOL and NTESOL 94

Table 4.15 Log-likelihood Ratios: The R1 collocates of *I* in LIVT, TESOL and NTESOL 95

Table 4.16 Log-Likelihood Ratios: *I was* and *was* 96

Table 4.17 Use of Past Tense Verbs in LIVT, TESOL and NTESOL 96

Table 4.18 Instance of the Verbs used by NTE07 97
Table 4.19 Log-Likelihood Ratios: *I would* and *would* 98
Table 4.20 Use of *I would* and *would* in LIVT, TESOL and NTESOL 98
Table 4.21 Use of *I would* and *would* in BASEah and COCAsp 99
Table 4.22 Log-Likelihood Ratios: *will* and *I will* 100
Table 4.23 Use of *will* and *I will* in LIVT, TESOL and NTESOL 101
Table 4.24 Instances of will used in LIVT and NTESOL 101
Table 4.25 Log-Likelihood Ratios: *I just* and *just* 102
Table 4.26 Log-Likelihood Ratios: *I think* and *think* 102
Table 4.27 Instances of *I think* used in LIVT 103
Table 4.28 Log-likelihood Ratios: *think that* 104
Table 4.29 R1 collocates of *think* in LIVT, TESOL and NTESOL 105
Table 4.30 Use of *think* and *think that* in BASEah and COCAsp 106
Table 4.31 Findings on the investigation of the use of *I* 107
Table 4.32 Log-Likelihood Ratios: *you* and 你(nǐ) in LIVT, TESOL and NTESOL 109
Table 4.33 L1 Collocates of *you* in LIVT, TESOL and NTESOL 110
Table 4.34 Log-Likelihood Ratios: The L1 collocates of *you* in LIVT, TESOL and NTESOL 111
Table 4.35 Log-Likelihood Ratios: *if you* and *if* 112
Table 4.36 Frequency of use of *if* and *if you* and other collocates of *if* in LIVT, TESOL and NTESOL 113
Table 4.34 Frequency of use of *if* ,*if you* and *if I* in BASEah and COCAsp 114
Table 4.35 Instances of *if you* in LIVT 114
Table 4.36 Patterns of *if you* in LIVT, TESOL and NTESOL 115
Table 4.37 Instances of *like if you* in LIVT 115
Table 4.38 Instances of like if you in COCAsp 116
Table 4.39 Log-Likelihood Ratios: *when you* and *when* 116
Table 4.40 Frequency of use of *when* and *when you* and other collocates of *when* in LIVT, TESOL and NTESOL 117
Table 4.41 Instances of *when you* in LIVT 117
Table 4.42 Patterns of *when you* in LIVT, TESOL and NTESOL 118
Table 4.43 Instances of * *when you* 118
Table 4.44 Instances of *that when you* in TE 119
Table 4.45 Instances of *and when you* in NTE 119
Table 4.46 Patterns of *when you* in BASEah and COCAsp 120
Table 4.47 Use of the collocations of *when you* in all groups 120
Table 4.48 Log-Likelihood Ratios: *what you* and *what* 121
Table 4.49 Frequency of use of *what you* and other collocates of *what* in LIVT, TESOL and NTESOL 121
Table 4.50 Patterns of *what you* in LIVT 122
Table 4.51 Instances of * *what you* 122
Table 4.52 Patterns of *what you* in BASEah and COCAsp 123

Table 4.53 R1 Collocates of *you* in LIVT, TESOL and NTESOL 124
Table 4.54 Log-Likelihood Ratios: The R1 collocates of *you* in LIVT, TESOL and NTESOL 125
Table 4.55 Use and Patterns of *you to* in LIVT, TESOL and NTESOL 126
Table 4.56 Instances of * *for you to* 127
Table 4.57 Instances of * *you to* in TE and NTE 128
Table 4.58 Use of Patterns of *you to* in LIVT, TESOL and NTESOL 128
Table 4.59 Patterns of *you to* in BASEah and COCAsp 129
Table 4.60 Findings on the investigation of the use of *you* 130
Table 4.61 Log-Likelihood Ratios: *it* and 它(tā) in LIVT, TESOL and NTESOL 132
Table 4.62 L1 collocates of *it* in LIVT, TESOL and NTESOL 133
Table 4.63 Log-Likelihood Ratios: The L1 collocates of *it* in LIVT, TESOL and NTESOL 134
Table 4.64 Use of *to it* in LIVT, TESOL and NTESOL 135
Table 4.65 Instances of *to it* in LIVT 135
Table 4.66 Use of *think* and *think it* in LIVT, TESOL and NTESOL 136
Table 4.67 Patterns of *think it* in LIVT, TESOL and NTESOL 136
Table 4.68 Use of *learn* and *learn it* in LIVT, TESOL and NTESOL 137
Table 4.69 Instances of *learn it* from NTESOL 138
Table 4.70 Use of *use*, *use it* and *how to use* in LIVT, TESOL and NTESOL 138
Table 4.71 Instances of *how to use* in TESOL 139
Table 4.72 R1 collocates of *it* in LIVT, TESOL and NTESOL 140
Table 4.73 Log-Likelihood Ratios: The R1 collocates of *it* in LIVT, TESOL and NTESOL 141
Table 4.74 Log-Likelihood Ratios: *it will* and *will* in LIVT, TESOL and NTESOL 141
Table 4.751 Use of *will* and *it will* in LIVT, TESOL and NTESOL 142
Table 4.76 Patterns of *it will* in LIVT, TESOL and NTESOL 143
Table 4.77 Patterns of *it will* in BASEah and COCAsp 143
Table 4.78 Findings on the investigation of the use of *it* 144
Table 4.79 Findings on the investigation of the use of pronouns *I*, *you* and *it* 147

Chapter 5

Table 5.1 L1 Collocates of 我(wǒ) in LIVT, TESOL and NTESOL 149
Table 5.2 Log-likelihood Ratios for L1 collocates of 我(wǒ) in LIVT, TESOL and NTESOL 150
Table 5.3 Use of 我(wǒ) 觉得(juéde) and *I think* in LIVT, TESOL and NTESOL 151
Table 5.4 The use of 就是(jiùshì) and 就是(jiùshì) 我(wǒ) in LIVT, TESOL and NTESOL 156
Table 5.5 R1 Collocates of 我(wǒ) in LIVT, TESOL and NTESOL 157
Table 5.6 Log-likelihood Ratios for R1 collocates of 我(wǒ) in LIVT, TESOL and NTESOL 158
Table 5.7 Log-likelihood Ratios: 我就(wǒ jiù) and 就(jiù) in LIVT, TESOL and NTESOL 160
Table 5.8 Collocates and patterns of use of 我就(wǒ jiù) and 就(jiù) in LIVT, TESOL and NTESOL 161
Table 5.9 Log-Likelihood Ratios: *I just, just,* 我就(wǒ jiù) *and* 就(jiù) 162
Table 5.10 Log-Likelihood Ratios: *I just, just,* 我就(wǒ jiù) and 就(jiù) 162

Table 5.11 Log-likelihood Ratios for 我还(wǒ hái) and 还(hái) in LIVT, TESOL and NTESOL 164
Table 5.12 Use of 我还(wǒ hái) and 还(hái) in LIVT, TESOL and NTESOL 165
Table 5.13 Instances of the use of 我还蛮喜欢(wǒ hái mán xǐhuɑn) in LIVT, TESOL and NTESOL 165
Table 5.14 Findings in the use of 我(wǒ) that Disagree with Hypothesis 3 167
Table 5.15 Findings in the use of 我(wǒ) that Disagree with Hypothesis 4 168
Table 5.16 L1 Collocates of 你(nǐ) in LIVT, TESOL and NTESOL 169
Table 5.17 Log-likelihood Ratios for L1 collocates of 你(nǐ) in LIVT, TESOL and NTESOL 170
Table 5.18 Log-likelihood Ratios: 时候(shíhòu) and 时候(shíhòu) 你(nǐ) in LIVT, TESOL and NTESOL 173
Table 5.19 Use of 时候(shíhou) 你(nǐ) and 你(nǐ) in LIVT, TESOL and NTESOL 174
Table 5.20 Instances of the use of 时候(shíhou) 你(nǐ) in LIVT, TESOL and NTESOL 175
Table 5.21 Use of 的时候(de shíhou) and 的时候(de shíhou)你(nǐ) in LIVT, TESOL and NTESOL 176
Table 5.22 Log-likelihood Ratios: 让(ràng) and 让你(ràng nǐ) in LIVT, TESOL and NTESOL 177
Table 5.23 Examples of the use of 让(ràng) 177
Table 5.24 Use of 让你(ràng nǐ) and 你(ràng) in LIVT, TESOL and NTESOL 178
Table 5.25 Examples of the use of 去(qù) 179
Table 5.26 Use of 让你(ràng nǐ), 让(ràng) and 去(qù) in LIVT, TESOL and NTESOL 179
Table 5.27 Use of 让(ràng)* *+verb*, 让(ràng)*去(qù) *+verb*, *let* *+ *verb* and *let * to +verb* in LIVT, TESOL and NTESOL 180
Table 5.28 R1 Collocates of 你(nǐ) in LIVT, TESOL and NTESOL 183
Table 5.29 Log-likelihood Ratios for R1 collocates of 你(nǐ) in LIVT, TESOL and NTESOL 184
Table 5.30 Log-likelihood Ratios: 你要(nǐ yào) in LIVT, TESOL and NTESOL 185
Table 5.31 Use of 要(yào) and 你要(nǐ yào) in LIVT, TESOL and NTESOL 186
Table 5.32 Use of 要(yào) in 你要(nǐ yào) in LIVT, TESOL and NTESOL 186
Table 5.33 Log-likelihood Ratios: 你说(nǐ shuō) in LIVT, TESOL and NTESOL 188
Table 5.34 Use of 你说(nǐ shuō) in LIVT, TESOL and NTESOL 188
Table 5.35 Examples of the four types of the use of 你说(nǐ shuō) in LIVT, TESOL and NTESOL 190
Table 5.36 The four types of the use of 你说(nǐ shuō) used by LIVT, TESOL and NTESOL 190
Table 5.37 Findings in the use of 你(nǐ)[you] that Disagree with Hypothesis 3 193
Table 5.38 Findings in the use of 你(nǐ)[you] that Disagree with Hypothesis 4 194
Table 5.39 L1 Collocates of 它(tā) in LIVT, TESOL and NTESOL 195
Table 5.40 Log-likelihood Ratios for L1 collocates of 它(tā) in LIVT, TESOL and NTESOL 196
Table 5.41 Use of 觉得(juéde) and 觉得(juéde) 它(tā) in LIVT, TESOL and NTESOL 197
Table 5.42 Use of 我(wǒ)觉得(juéde) and 我(wǒ)觉得(juéde)它(tā) in LIVT, TESOL and NTESOL 198
Table 5.43 Examples of another three types of the use of 觉得(juéde) 它(tā) in LIVT, TESOL and NTESOL 199
Table 5.44 Use of 了解(liǎojiě) and 了解(liǎojiě)它(tā) in LIVT, TESOL and NTESOL 202
Table 5.45 R1 Collocates of 它(tā) in LIVT, TESOL and NTESOL 203
Table 5.46 Log-likelihood Ratios for R1 collocates of 它(tā) in LIVT, TESOL and NTESOL 204
Table 5.47 Frequency and Log-likelihood Ratios: 它那(tā nà) in LIVT, TESOL and NTESOL 205
Table 5.48 Examples of the use of 它那(tā nà) 206

Table 5.49 Use of the patterns *pronoun+ 那(nà) [(of) it, that] / pronoun + 的(de) +那(nà) [of it, that]* 207
Table 5.50 Findings in the use of 它(tā) [it] that disagree with Hypothesis 3 209
Table 5.51 Findings in the use of 它(tā) [it] that disagree with Hypothesis 4 210
Table 5.52 Summary of the findings that disagree with Hypothesis 3 –Discoursal Preferences 213
Table 5.53 Summary of the findings that disagree with Hypothesis 3 –Lexico-grammatical Differences 214
Table 5.54 Summary of the findings that disagree with Hypothesis 4 –Discoursal Preferences 214
Table 5.55 Summary of the findings that disagree with Hypothesis 4 –Lexico-grammatical Differences 215

Chapter 6

Table 6.1 Frequencies of *is* in LIVT, TESOL and NTESOL 216
Table 6.2 Frequencies of 是(shì) in LIVT, TESOL and NTESOL 216
Table 6.3 Log-Likelihood Ratios: *BE-verbs* in LIVT, TESOL and NTESOL – L1 Collocates 218
Table 6.4 Use of Aux. Be-verbs in LIVT, TESOL and NTESOL 219
Table 6.5 Log-Likelihood Ratios: *BE-verbs* in LIVT, TESOL and NTESOL – R1 Collocates 220
Table 6.6 L1 collocates of *is* in LIVT, TESOL and NTESOL 221
Table 6.7 Log-likelihood Ratios: Collocations of *is* – L1 Collocates 222
Table 6.8 L1 collocates of *'s* in LIVT, TESOL and NTESOL 222
Table 6.9 Log-likelihood Ratios: Collocations of *'s* – L1 collocates 223
Table 6.10 Log-likelihood Ratios: Collocations of *was* – L1 collocates 224
Table 6.11 L1 collocates of *was* in LIVT, TESOL and NTESOL 224
Table 6.12 Log-likelihood Ratios: *it is* and *it's* 225
Table 6.13 Use of *it is* and *it's* in LIVT, TESOL and NTESOL 226
Table 6.14 Use of *it is because* and *it's because* in LIVT, TESOL and NTESOL 226
Table 6.15 Use of *I think it is because/it's because* in LIVT, TESOL and NTESOL 227
Table 6.16 Instances of the use of *I think it is because/it's because* in LIVT, TESOL and NTESOL 228
Table 6.17 Use of *that is* and *that's* in LIVT, TESOL and NTESOL 229
Table 6.18 Use of *that is* and *that's* in BASEah and COCAsp 230
Table 6.19 R1 collocates of *is* in LIVT, TESOL and NTESOL 232
Table 6.20 R1 collocates of *'s* in LIVT, TESOL and NTESOL 233
Table 6.21 Log-likelihood Ratios: Collocations of *is* – R1 Collocates 233
Table 6.22 Log-likelihood Ratios: Collocations of *'s* – R1 Collocates 234
Table 6.23 Log-likelihood Ratios: Collocations of *was* – R1 Collocates 234
Table 6.24 Examples of the use of *is that* in LIVT, TESOL and NTESOL 236
Table 6.25 Use of *is that* in LIVT, TESOL and NTESOL 236
Table 6.26 Log-Likelihood Ratios: *NP+ is that + [S]* 236
Table 6.27 Log-Likelihood Ratios: *quite* and *is/'s quite* 238
Table 6.28 Use of *quite* and *is quite* in LIVT, TESOL and NTESOL 238
Table 6.29 Use of adjectives with *is('s) quite* in LIVT, TESOL and NTESOL 239
Table 6.30 Use of adjectives with *is('s) quite* in BASEah and COCAsp 239

Table 6.31 Log-Likelihood Ratios: *pretty* and *is('s) pretty* 241
Table 6.32 Use of *pretty* and *is('s) pretty* in LIVT, TESOL and NTESOL 241
Table 6.33 Use of *pretty* and *is('s) pretty* in BASEah and COCAsp 242
Table 6.34 Use of adjectives with *is('s) pretty* in LIVT, TESOL, BASEah and COCAsp 243
Table 6.35 Participants using *is('s) quite* and *is('s) pretty* in LIVT, TESOL and NTESOL 244
Table 6.36 Findings in the use of is that support Hypothesis 1 245
Table 6.37 Findings in the use of is that support Hypothesis 2 246
Table 6.38 Log-Likelihood Ratios: 是(shì) in LIVT, TESOL and NTESOL 248
Table 6.39 Use of *是 (shì)in LIVT, TESOL and NTESOL 249
Table 6.40 Log-Likelihood Ratios: *是(shì) in LIVT, TESOL and NTESOL 250
Table 6.41 L1 Collocates of 是(shì) in LIVT, TESOL and NTESOL 251
Table 6.42 Log-Likelihood Ratios: 是(shì) in LIVT, TESOL and NTESOL – L1 Collocates 251
Table 6.43 Log-Likelihood Ratios: 我是(wǒ shì)and 是(shì) 252
Table 6.44 Use of 是(shì) and 我是(wǒ shì) in LIVT, TESOL and NTESOL 253
Table 6.45 Use of 觉得(juéde), 是(shì)觉得(juéde), 我(wǒ)觉得(juéde) and 我(wǒ)是(shì)觉得(juéde) in LIVT, TE and NTE 256
Table 6.46 Use of 我是(wǒ shì) in LIVT, TESOL and NTESOL 257
Table 6.47 Log-Likelihood Ratios: 它是(tā shì) 259
Table 6.48 Use of 是(shì) and 它是(tā shì) in LIVT, TESOL and NTESOL (1) 259
Table 6.49 Use of 是(shì) and 它是(tā shì) in LIVT, TESOL and NTESOL (2) 262
Table 6.50 R1 Collocates of 是(shì) in LIVT, TESOL and NTESOL 264
Table 6.51 Log-Likelihood Ratios: 是(shì) in LIVT, TESOL and NTESOL – R1 Collocates 265
Table 6.52 Log-Likelihood Ratios: 是你(shì nǐ) 266
Table 6.53 Use of 是(shì) and 是你(shì nǐ) in LIVT, TESOL and NTESOL 266
Table 6.54 Use of 是(shì) in 是你(shì nǐ) in LIVT, TESOL and NTESOL 268
Table 6.55 Log-Likelihood Ratios: 是我(shì wǒ) 269
Table 6.56 Use of 是(shì) and 是我(shì wǒ) in LIVT, TESOL and NTESOL 269
Table 6.57 Use of 是(shì) in 是我(shì wǒ) in LIVT, TESOL and NTESOL 270
Table 6.58 Log-Likelihood Ratios: 是说(shì shuō) 272
Table 6.59 Use of 是(shì) and 是说(shì shuō) in LIVT, TESOL and NTESOL 272
Table 6.60 Patterns of 是说(shì shuō) in LIVT, TESOL and NTESOL 277
Table 6.61 Findings in the use of 是 (shì) that disagree with Hypothesis 3 279
Table 6.62 Findings in the use of 是 (shì) that disagree with Hypothesis 4 280
Table 6.63 Summary of the findings of the use of *is* that support Hypothesis 1 281
Table 6.64 Summary of the findings of the use of *is* that support Hypothesis 2 282
Table 6.65 Summary of the findings that Disagree with Hypothesis 3 283
Table 6.66 Summary of the findings that Disagree with Hypothesis 4 284

Chapter 7

Table 7.1 Frequencies of *have* in LIVT, TESOL and NTESOL 287

Table 7.2 Frequencies of 有(yǒu) in LIVT, TESOL and NTESOL 287
Table 7.3 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL 288
Table 7.4 Frequencies of *HAVE-Verbs* in LIVT, TESOL and NTESOL 289
Table 7.5 Use of *HAVE-Verbs* in LIVT, TESOL and NTESOL 289
Table 7.6 Log-Likelihood Ratios: *'ve* in LIVT, TESOL and NTESOL 289
Table7.7 L1 Collocates of *have* in LIVT, TESOL and NTESOL 291
Table 7.8 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL – L1 Collocates 292
Table 7.9 Log-Likelihood Ratios: *I've* in LIVT, TESOL and NTESOL 292
Table 7.10 Use of *'ve* in LIVT, TESOL and NTESOL 292
Table 7.11 Use of *I have* and *I've* in LIVT, TESOL and NTESOL 293
Table 7.12 Instances of *'ve* in NTESOL 294
Table 7.13 Use of *I have* and *I've* in BASEah and COCAsp 294
Table 7.14 Instances of *I have*/*I've* in NTE 295
Table 7.15 Use of *I have* in LIVT, TESOL and NTESOL 296
Table 7.16 Use of *I + aux. have* in LIVT, TESOL and NTESOL 297
Table 7.17 R1 Collocates of *have* in LIVT, TESOL and NTESOL 298
Table 7.18 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL – R1 Collocates 299
Table 7.19 Use of *have* and *ha* to* in LIVT, TESOL and NTESOL 300
Table 7.20 Use of *have* and *ha* to* in LIVT, TESOL and NTESOL 301
Table 7.21 Log-Likelihood Ratios: *VH+V/PP* in LIVT, TESOL and NTESOL 302
Table 7.22 Findings in the use of *have* that support Hypothesis 1 303
Table 7.23 Findings in the use of *have* that support Hypothesis 2 304
Table 7.24 L1 Collocates of 有(yǒu) in LIVT, TESOL and NTESOL 306
Table 7.25 Log-Likelihood Ratios: 有(yǒu) in LIVT, TESOL and NTESOL – L1 collocates 306
Table 7.26 Log-Likelihood Ratios: 还有(háiyǒu) [still have/exist] in LIVT, TESOL and NTESOL 307
Table 7.27 Frequencies of the use of 还有(hái yǒu) [in addition] in LIVT, TESOL and NTESOL 308
Table 7.28 Examples of the use of 还有(hái yǒu) [in addition] 308
Table 7.29 The use of the word 有(yǒu) in 还有(hái yǒu) in LIVT, TESOL and NTESOL 309
Table 7.30 Log-Likelihood Ratios: 还是有(hái shì yǒu) in LIVT, TESOL and NTESOL 311
Table 7.31 Use of 还是有(hái shì yǒu) in LIVT, TESOL and NTESOL 311
Table 7.32 Examples of the use of 还是有(hái shì yǒu) [still SHI(to be) have] 312
Table 7.33 R1 Collocates of 有(yǒu) in LIVT, TESOL and NTESOL 314
Table 7.34 Log-Likelihood Ratios: 有(yǒu) in LIVT, TESOL and NTESOL – R1 collocates 314
Table 7.35 Use of 有(yǒu) + *NOUN* in LIVT, TESOL and NTESOL 318
Table 7.36 Log-Likelihood Ratios: 有(yǒu) + *NOUN* in LIVT, TESOL and NTESOL 318
Table 7.37 Use of 有(yǒu)+ *VERB* in LIVT, TESOL and NTESOL 319
Table 7.38 Log-Likelihood Ratios: 有(yǒu)+ *VERB* in LIVT, TESOL and NTESOL 319
Table 7.39 Use of 参加(cānjiā) in LIVT, TESOL and NTESOL 322

Table 7.40 Findings in the use of *有 (yǒu) [have/exist]* that disagree with Hypothesis 3 325
Table 7.41 Findings in the use of *有 (yǒu) [have/exist]* that disagree with Hypothesis 4 326
Table 7.42 Summary of the findings of the use of *have* that support Hypothesis 1 327
Table 7.43 Summary of the findings of the use of *have* that support Hypothesis 2 327
Table 7.44 Summary of the findings of the use of 有(yǒu) that Disagree with Hypothesis 3 328
Table 7.45 Summary of the findings of the use of 有(yǒu) that Disagree with Hypothesis 4 328

Chapter 1 Introduction

1.1 The motivation behind this research

The idea behind this research originated from my personal experience of studying abroad. In 2010 I came to the UK to pursue an MA programme. It was my first time in an English-speaking country and my first time living abroad for an extended period. Although I had received formal compulsory English education at junior and senior high school for over six years and another four years of training in the teaching of English as a second/foreign language during my Bachelor's degree in Taiwan, I still struggled to use English in my academic studies and in daily life when abroad. I had trouble finding the right expressions to accurately express my experience in English. Moreover, apart from my lack of knowledge of certain English expressions, I came to realise that my English, especially my spoken English, had an 'oddness' about it that distinguished me from the native speakers, even when I was using the expressions that I had been taught correctly.

When native English speakers mention the 'oddness' of Taiwanese speakers and other non-native speakers, the accent is the first characteristic mentioned. However, in my personal experience, the accent is not the main issue. Moreover, since English is now recognised to be a collection of so-called *World Englishes* that encompass the broad varieties used by different regional and ethnic groups (Kachru, 1982, 2006), accent is no longer an important signifier of 'nativeness'. Indeed the accents of English native speakers differ according to region or ethnic group and this phenomenon is not unique to English speakers. For instance, my Taiwanese Mandarin accent is noticeable to the Mandarin speakers I encounter from other regions. However, there is no issue regarding the fact that my Mandarin differs from that of a native Chinese person's Mandarin.

I have noticed (and believe many EFL/ESL teachers would agree) that the main reason my spoken English sounds 'odd' is my choice of lexis; that is, it is a direct result of the words or phrases I select and the grammatical structures linked to those words and phrases. Sometimes, the sentences I produce are ungrammatical; more often, the sentences are grammatical but would not be used by native speakers or, if they did use them, it would be in other, very specific contexts.

The phenomenon of *fossilisation*

One explanation for this 'unnatural/odd' language use might lie with the phenomenon of *fossilisation* that has been defined and discussed by many Second Language Acquisition (SLA) researchers (Selinker, 1972; Birdsong, 1992; Schachter, 1990, 1996; Schwartz, 1998; Lightbown & Spada, 1999; MacWhinney, 2001; Han, 2004.) As Han (2004: 13) puts it:

> *This phenomenon of non-progression of learning despite continuous exposure to input, adequate motivation to learn, and sufficient opportunity of practice - generally referred to in the literature as 'fossilization' – became a central concern for SLA researchers.* (Han, 2004: 13)

The concept of *fossilisation* was explained by Selinker (1972) in his Interlanguage Hypothesis. The term *interlanguage* is defined as the separate linguistic system possessed by adult second language learners. The interlanguage differs from both the second language learners' native language system and their target language/second language system. Mastering a target language requires one's interlanguage to become less similar to one's native language system and more similar to one's target language system. The ultimate goal of a second language

learner is to develop an interlanguage system that resembles the target language system. The phenomenon of fossilisation is central to the notion of interlanguage. Fossilisation happens when a learner's interlanguage stops developing, at which point fossilised development seems permanent. Selinker (1972: 22) implies that this phenomenon involves a cognitive mechanism, called the *fossilisation mechanism*, that interrelates with learners' linguistic performance.

Selinker's (1972) view of the relationship between SLA and *fossilisation* has expanded over the years and his final position is summarised in Han (2004) as follows:

> *[T]he ultimate attainment of adult L2 acquisition is a fossilized interlanguage; fossilization is inevitable; and no adult L2 learner would ever be able to pass for native in all contexts.* (Han, 2004:15)

Selinker's (1972) view is that: (1) the language used by a non-native speaker is a fossilised interlanguage; (2) non-native speakers have a limited cognitive competence to achieve nativelikeness in the target language; and, (3) this inability is inevitable. The fossilised interlanguage is either restricted only to the L2 learners' language or every one's language is a 'fossilised' language. Whether there is a limit to L2 learners' cognitive capacity is still a point of debate. While I agree that the best way to research this kind of phenomenon is through the study of people's psychological and cognitive processes, confirmation of this requires further empirical evidence. As Jung (2002: 16; quoted by Han, 2004: 22) states:

> *To make any decisive claims [on fossili*z*ation] … it would be necessary to demonstrate that the fossilized item in question has completely ceased developing towards the L2 norm.*

Nevertheless, the long-lasting debate about *fossilisation* (and also about *interlanguage* since the two concepts are interlinked) indicates that the phenomenon observed from my personal experience is a topic worth studying.

A seldom explored assumption in SLA

Another reason why I undertook this research was to investigate a seldom explored assumption in SLA, namely that, "the only way to achieve real fluency in a foreign language is to go to a place where it is spoken" (Ferguson, 1995: xii; in Freed, 1995). Ferguson (1995) further states that this is an SLA 'myth', which has rarely been examined systematically. Freed (1995) adds that existing studies only cover a limited area in terms of the perspective of linguistic uses:

> *Surprisingly […], there have been relatively few empirical studies which have addressed, in a carefully-controlled and in-depth manner, the specific question of the linguistic impact of study abroad. While the general benefits of study abroad have been widely described in a number of publications […], these evaluative reports tend to deal with issues such as preparation for the study abroad experience, program assessment, student evaluations, general policy issues, and overall benefits that result from a study abroad experience. […], they contain little empirically-based research which describes or analyses the impact of a study abroad experience on the language skills of students who have been abroad.* (Freed, 1995: 5-6)

Freed (1995: 6) continues, stating that SLA studies regarding this issue have delivered conflicting findings. She categorises these studies into two types of study. The first type of study investigates the relationship between out-of-class contact and language acquisition in a

language immersing environment. Some research (Bialystok, 1978; Rubin, 1975; Seliger, 1977 and Stern, 1983) shows that learners who try the hardest to use the target language in out-of-class situations will be those who make the best progress. However, subsequent similar research has produced contradictory evidence. These studies (Day, 1985; DeKeyser, 1986; Freed, 1990; Krashen and Seliger, 1976; Krashen et al., 1974; Spada, 1985, 1986) suggest that out-of-class immersion of language does not guarantee good progress.

The second kind of SLA study focuses on evaluating access to Universal Grammar (UG) by two sorts of learners: those who receive language learning in a simple classroom-based environment and those who experience immersion in the target language community. Although these studies concern UG, their primary focus is not on the assessment of SL learners' linguistic performance. Rather they emphasise the critical period effect in SLA. In a study of second language learners' access to UG, Felix and Weigl (1991) found that language classroom teaching seems to systematically restrain learners' access to UG and the process of acquisition. In contrast, a study of two groups of Chinese students, undertaken by White and Juffs (1998), showed that the group with formal classroom teaching performed better on tests that required grammatical judgment than the group immersed in an environment where English was spoken (Freed, 1995:6-7).

When assessing these studies, Freed (1995) concluded that the question of whether immersion is helpful or not remains unsettled. She also points out that a major limitation of some studies was that they used only test scores to measure linguistic skills. Whether such scores can truly reveal qualitative changes in learners' language proficiency warrants further investigation (1995:9). She suggests the size and duration of the research, the absence of control groups, and the use of test scores as the evaluation means are limitations which need to be resolved (1995: 16).

Moreover, as Freed notes:

> *With a few notable exceptions, the findings from the studies to date tell us little about actual language use and serve, therefore, as preliminary explorations of the topic. No study has yet described a range of linguistic variables (phonologic, syntactic and/or semantic) sociolinguistic and discourse features, that may be influenced as a result of a study abroad experience* (Freed, 1995: 16).

She also points out the lack of research on linguistic variables, sociolinguistic and discourse features. This type of research "provides a wealth of opportunity for research on the linguistic impact of study abroad, as well as on the cognitive and social aspects of the second language acquisition process" (Freed, 1995:6).

My personal experience, in conjunction with the phenomenon of *fossilisation* and the SLA assumption of the positive value of immersion, has therefore motivated me to carry out this research.

1.2 The potential value of this research

For the reasons discussed above, I conducted an investigation into the language use of EFL learners whose language learning experiences differ based on several controlled variables. The study encompassed three different scenarios: (1) EFL learners receiving their postgraduate education abroad; (2) EFL learners receiving English-based postgraduate professional training in their country of origin; (3) EFL learners receiving postgraduate education which is not English-based in their country of origin (as the control group). The focus of this thesis is purely the linguistic performance of these EFL learners. I applied the lexical priming theory (Hoey, 2005) in my investigation of the participants' language use. This is a theory which views lexical items as the primary elements in people's language use and uses the methodology of corpus linguistics. Using the research design described in detail in Chapter 3, the aim of this study was to provide some insight into how EFL learners' linguistic performance is influenced by these different variables (immersion, explicit linguistic input in a non-immersion context, no explicit input in a non-immersion context) in response to the discussion of SLA assumption.

The contribution of this study is threefold. This research will potentially fill the gap identified by Freed (1995) regarding the effect of studying abroad on EFL learners' linguistic use. It will potentially provide more information regarding whether lexical priming theory can be used to describe people's language use. Finally, it will also potentially provide some insight into our understanding of the phenomenon of *fossilisation* in SLA.

1.3 The structure of this thesis

In the next chapter I will discuss the theoretical background to this study, specifically the role of input second language acquisition, the different perspectives on language acquisition and the way people view language use. I will give my research hypotheses at the end of Chapter 2. In Chapter 3, I will provide a thorough description of the methodology used in this study, including the criteria used in participant selection, the process of language material collection, a detailed description of the participants of the study, the methods and tools used for analysable corpus compilation and the method of data analysis. Chapters 4 to 7 deal with the use of different lexical items across the three groups/corpora. Chapters 4 and 5 concern the use of personal pronouns *I*, *you* and *it* in English and 我 [I], 你 [you] and 它 [it] in Mandarin Chinese. Chapter 6 concerns the use of the functional verbs *is* in English and 是 [to be] in Mandarin Chinese. Chapter 7 concerns the use of the verb *have* in English and 有 [to have/to exist] in Mandarin Chinese. All of the investigations into these English and Mandarin lexical items will focus on their collocations and the frequently occurring word clusters they are associated with. The overall conclusions of this research will be given in the final chapter, along with suggestions for future research.

Chapter 2 Theoretical Background

Chapter 1 introduced the two phenomena that motivated me to conduct this study. The first being the unnatural use of language by EFL learners, whereby ungrammatical, or grammatical but odd use of language prevents them from becoming as fluent in English as a native speaker. This is apparent even in EFL learners who have spent many years learning English, some even living in an English environment; they exhibit *fossilisation* in their use of English. This relates to the second issue of interest; i.e. whether studying abroad (immersion in the target language context) genuinely improves language performance, allowing EFL learners to become native-like EFL speakers. Exploring theories associated with unnatural language use, fossilisation, and the potential effects of immersion in the target language, this chapter describes and highlights some of the concepts researchers have associated with second language acquisition (SLA), to investigate why unnatural use of English persists in EFL learners' speech and writing.

2.1 A brief review of SLA theories

Chapter 1 briefly discussed *fossilisation* in language learning as it relates to the unnaturalness of EFL learners' language use, i.e. the presence of non-idiomatic combinations of words and misuse of language in discourse. On occasion, unnatural language takes the form of an unusual choice of words within an ungrammatical sentence. This type of language misuse arises when EFL learners have not fully understood the target language, and are still in the process of mastering it. There are also multiple examples of unnatural language use within grammatical sentences. That is, native listeners can readily comprehend what an L2 speaker's sentence is intended to express, but might perceive a peculiarity in the lexis chosen. This situation is commonplace among EFL learners who have been learning English for an

extended period of time and have achieved a good level of accuracy in formal English tests. Some of these EFL learners have only learned English in an English classroom setting, and have not been exposed to English in their daily lives. Arguably this situation arises because what is taught in the English classroom is too limited to cover the scope of English language use required. This might the oddities present in EFL learners' discourse when they encounter a situation not modelled in the classroom. However, idiosyncratic language use can also be observed among those EFL learners who have studied English while living in an English speaking environment. This has prompted researchers to consider how EFL learners acquire their second/foreign language.

Language use is a social phenomenon that allows people to communicate (Scovel, 2001). Therefore, it is crucial to understand how social cues inform language acquisition. Behaviourists, for example, suggest people learn language through a process of mimicry and memorisation, developing language in response to positive or negative reinforcements arising from the immediate environment (Skinner, 1957; Brook, 1960; Lado, 1964). Alternatively, innatists claim that people are born with a pre-programmed linguistic mechanism in place (Chomsky, 1956, 1986), termed Universal Grammar (UG). According to this view, language learning arises naturally as a consequence of human development, and while people superficially use language differently, the same neurological structures are responsible for language acquisition and use.

The social interactionists claim that language is learned through interaction, and thus the phenomenon whereby people learn a language arise from social interaction. Interaction with other people can enable learners to advance from their current level of linguistic development to a higher level of linguistic development (Vygotsky, 1962, 1978; Long, 1983, 1996). A further proposal is set out by psycholinguists and cognative psychologists, who claim that

linguistic knowledge is acquired as part of the information-processing model of human learning and performance. According to this view, an information-processing model is a skilled learning mechanism well suited to building up a variety of skills, not merely linguistic skills (Anderson and Spellman, 1995; DeKeyser, 1998, 2001, 2007). Language learning is viewed to begin with declarative knowledge, which people consciously possess, such as grammar rules, and that this declarative knowledge transforms into procedural knowledge, which people draw on but are unaware of knowing. Eventually this knowledge of language becomes automatic and fluency is achieved. This transformation happens when use of knowledge occurs frequently and continuously (Lightbown & Spada, 2013: 109-20).

It is crucial to note that behaviourists and social interactionists are not wholly divergent in their views. They agree that language learners' engagement with their environment is the most crucial aspect of language learning. The former suggest people imitate the linguistic input they receive and reinforce or amend their language according to responses to their output. Meanwhile, the latter argue that people learn a language through negotiating meaning with others. During this process, previously acquired language is stabilised and previously unknown language is assimilated from interlocutors. Both consider how the outer world affects language acquisition, viewing it as more significant than the mental processes involved in learning a language. In contrast, innatists focus on the neurological mechanisms involved when people acquire language. They claim language learning is reliant on an autonomous linguistic device that underpins natural development, viewing the influence of the environment as relatively insignificant by contrast.

Arguably, there are similarities between the perspectives of innatists and psycholinguists/cognitive psychologists. Both focus on what happens in people's minds when they acquire a language as their predominant concern when resolving SLA theories. However,

there is a fundamental difference between them. Innatists believe there is a pre-wired language system in the human brain, and that linguistic stimuli in the environment trigger adjustments (principles and parameters) to this language system, encouraging the deep structure of language (the inborn form) to create a surface structure of language (the form used in the environment) to facilitate communication. Universal Grammar-based researchers and some cognitive theorists (Towell and Hawkins, 1994) consider language "a separate module in the mind, distinct from other aspects of cognition" (Mitchell and Myles, 2004: 95). The language knowledge we gain is processed uniquely. However, some cognitive theorists view the ability to learn a language as just one aspect of general cognitive ability (Ellis, 2003, MacWhinney, 1999 and Tomasello, 1992). From this perspective, the way people process linguistic information is likened to the processing of other types of information. Those theorists who view language processing as a special cognitive ability set out specific processing approaches, whereas, theorists who perceive language processing as similar to other types of information processing follow emergentist or constructionist approaches (Mitchell and Myles, 2004: 97). We will return to this conflict in section 2.2.

Human beings have an innate ability to learn language, although how this cognitive competence operates is a subject of debate. The research on language learning theories cites the significant involvement of environmental factors as key. In particular, we must emphasise the conditions of first language acquisition (FLA) and SLA (SLA) differ. For example, L2 (second language) learners usually have mature cognition when they first come to learn the target language, because they have already mastered at least one language. Furthermore, L2 acquirers can have very different motivations for learning a language from L1 (first language) acquirers. They already have the ability to communicate using language, while L1 acquirers are must build and master linguistic capacity. In this case, evidence concerning the two kinds of acquirers ought not to be conflated, despite sharing certain elements in common. This

thesis concerns the potential influence of environmental factors on EFL learners (L2 acquirers), and thus requires that we focus on the perspectives provided by researchers, who deem environmental factors to have a significant influence on language learning. It aims to answer the following questions, amongst others: How does the environment affect EFL learners' linguistic output? To what extent are language learning outcomes affected or reshaped under certain conditions?

2.1.1 Input Hypothesis, Output Hypothesis and Interaction Hypothesis

The most renowned theories positing the influence of the environment on language learning are: Krashen's Input hypothesis (1982, 1985, 1998), Swain's Output hypothesis (1985, 1995, 2005), and Long's Interaction hypothesis (1981, 1983, 1996).

Input Hypothesis

The Input hypothesis proposed by Krashen (1976, 1982, 1985, 2003), and subsequently renamed the Comprehension hypothesis, states that input is the most important element among the multiple environmental factors known to influence SLA. The Input hypothesis is one of five hypotheses set out in Krashen's theory of SLA. As Krashen himself states, it is at the heart of SLA theory (Krashen, 1982, 1985). The hypothesis asserts that humans acquire language through exposure to input that is a little beyond their current level of acquired competence (*i*+1) (Krashen and Terrell, 1983: 32): "[H]umans acquire language in only one way – by understanding messages, or by receiving 'comprehensible input'" (Krashen, 1985: 2). The messages to which learners are exposed may contain grammatical rules yet to be assimilated by L2 acquirers. In order to understand language that contains unacquired rules, learners must examine the context in which the new language appears. This context can

include extra-linguistic information, real-world knowledge, and previously learnt linguistic knowledge. Comprehensible input is the all-important environmental element that functions in combination with an "internal language processor" (Krashen, 1985). This reveals that social interactionists do not deny the existence of an innate, pre-wired mechanism in the brain. That is, they accept there is a black box, but have no desire to investigate what is inside it. With comprehensible input as the external environmental ingredient, and an internal mental language processor, Krashen (1982, 1985) states that both first and second languages can be eventually acquired. He further emphasises that it is a learners' willingness to process comprehensible input that is important. This willingness arises from attitude and is linked to motivation; this is discussed as the Affective Filter hypothesis. He suggests that people acquire a language only when they obtain comprehensible input, while their affective filters are low enough to permit input to enter their internal mental language processor. When the filter is 'down' and sufficient comprehensible input is presented, acquisition occurs (Krashen, 1985: 4).

To clarify how this works in practice, Krashen identifies several inferences arising from the Input hypothesis. One is that if sufficient comprehensible input is provided, the grammar/structures which learners need to acquire will emerge in the input. This means, there is no need to deliberately focus on a certain grammatical forms. This view suggests deliberately learning grammar rules is no guarantee of the successful acquisition of L2. Evidence for comprehensible input as the crucial factor in SLA is provided by Hammond (1988). Hammond (1988) reported that students who received comprehensible input in a school setting performed slightly better on grammar tests than those who did not. Empirical studies by Winitz (1996) and Nikolov and Krashen (1997) have demonstrated that deliberate teaching of grammar rules is less successful than teaching with comprehensible input. In their research, students who learned through exposure to comprehensible input performed better in

terms of the accuracy and fluency of the target language than those taught grammar rules. DeKeyser and Sokalski (1996) also conducted an experiment to evaluate the relationship between input, output and learning. They discovered that input and output practice influences different skills: input benefits learners' comprehension and output supports production. Although this evidence does not explain the phenomenon of unnaturalness in EFL learners' English use, it does demonstrate that input is critical in SLA. Moreover, provision of sufficient comprehensible input can, to some degree, influence EFL learners' language acquisition.

Input from the language classroom undoubtedly benefits students' SLA, and input outside the language classroom (see chapter one) results in better progress according to Bialystok (1978), Rubin (1975), Seliger (1977), and Stern (1983) (although other studies (Day, 1985; DeKeyser, 1986; Freed, 1990; Krashen and Seliger, 1976; Spada, 1986; Spada and Lightbown, 1993) find little or no relationship between exposure and progress). Krashen (1982, 1985) argues that it is when exposure to a second language typically includes comprehensible input, that a positive relationship is most likely to be seen. When exposure does not include comprehensible input, its value is very limited. Thus, it is apparent that exposure to the target language/Immersion in the environment of the target language does not equate to receiving comprehensible input. There are several elements to consider here: the learner's physical factors, e.g. the ability to catch the input by hearing/reading; the learner's cognitive factors, e.g. the learner's understanding of real world knowledge and capacity for learning; the learner's mental factors, e.g. the learner's attitude towards or willingness to process the input/environment.

If immersion in the target language is to prove effective, according to Krashen's hypothesis, the language immersion environment must contain abundant comprehensible

input and comprise a positive environment to negotiate the Affective Filter. The nature of the type of exposure that takes place in the environment is also relevant in SLA. Comprehensible input can arise during formal classroom teaching or in informal daily conversation. That is to say, the "formal and informal environment contribute to second language competence in different ways or rather, to different aspects of second language competence" (Krashen, 1988: 47). To support SLA the informal environment needs to be communication intensive, engaging and involving learners directly (Krashen, 1988). Exposure to input and immersion in the target language environment are central to SLA. They provide an important ingredient, comprehensible input, to benefit L2 learner's language acquisition. Therefore, this study investigates how exposure to input and language immersion influence L2 learners' language use in detail.

Interaction Hypothesis

While Krashen stressed the importance of comprehensible input in SLA, Long (1981, 1983) advanced conceptualisation of the role of comprehensible input, claiming that interaction between language users should be emphasised more consistently. This hypothesis is termed the Interaction hypothesis (Long, 1981, 1983), and is based on the belief that input is not one-directional. Thus, when learners engage in conversations and try to negotiate meaning, the input they receive changes in response. The input may be queried, recycled and paraphrased as the interlocutors employ communication strategies, such as repetition, confirmation and comprehension checks, or clarification requests, to ensure the language is comprehensible to the language learners. This type of interaction enriches the language input, enhancing the learners' comprehension; this explains why the role of interaction is deemed crucial in the context of SLA. Long (1996) further emphasises the connections between

comprehensible input and the linguistic environment in which language learners engage, acknowledging both are crucial to the development of a second language.

> *It is proposed that environmental contributions to acquisition are mediated by selective attention and the learner's developing L2 processing capacity, and that these resources are brought together most usefully, although not exclusively, during negotiation for meaning.*
>
> (Long, 1996: 414)

The Input Hypothesis as proposed by Krashen suggests exposure to comprehensible input is critical to SLA. Long considers interaction (negotiation for meaning) to be as important as the input process. In the above discussion of the circumstances of immersion in the environment of the target language/studying abroad (in the section of Input Hypothesis), it was explained that exposure to the target language/immersion is not necessarily the same as comprehensible input, as cognitive, mental, physical, and experiential factors might also influence whether input is comprehensible. Following on from this, Long's Interaction Hypothesis elaborates further, explaining that learners do not merely receive comprehensible input from the environment. It is produced mutually by the learners themselves, and the interlocutors around them. That is, comprehensible input does not merely describe the presentation of linguistic materials as provided by the environment, but is also a product created through collaboration between learners and the environment (other people).

Output Hypothesis

The notion that comprehensible input is sufficient for learners to achieve second language development was challenged by Swain (1985, 1995, 2005). She observed immersion

students exposed to rich comprehensible French input who had attained the level of native speakers in French. However, she noticed that the productive ability of these students was not synchronised with their receptive ability, and consequently argued that when language learners comprehend second language texts, they focus on semantic processing rather than processing the input as a whole. Therefore, Swain claims, comprehensible input is insufficient for learners to master a second language. She argues that it is rather second language production (output) that guarantees successful SLA. Only when students succeed in processing language for production (output), can the development of syntax and morphology of the second language truly be achieved. This hypothesis is referred to as the Output Hypothesis (Swain, 1985, 1995, 2005).

It is difficult to deny the role of output (practice in producing language) in language development as regards achieving fluency and mastery. However, Swain's Output Hypothesis makes some claims that are beyond the domain of 'practice makes perfect'. She suggests output plays an important role in the development of the interlanguage system, not simply as a means of improving language skills. Swain claims that through the practice of producing a language, learners will become aware of gaps and limitations in their current second language system, which prevent them from conveying their ideas efficiently and accurately. The need to resolve these gaps will lead them to discuss and analyse the language explicitly, and offer opportunities to experiment with new structures and forms (Swain, 1995).

The awareness of 'gaps' between language users' conversation and language used in output are seen as key to successful SLA in Swain's Output Hypothesis (Swain, 1995, Swain and Lapkin, 1995, 1998). Several researchers have conducted research in response to her theory. Ellis and He (1999) conducted an experiment relating to vocabulary learning, and discovered that learners required to produce output (giving instructions to others) performed

better than learners who only received instructions. The evidence suggests that practice producing output and modifications to output might significantly advance language acquisition. Therefore, if language learners produce language in use, the probability that they will become aware of the inadequacy of their current competence in language is greater, which then leads to developments in their language learning. Simple exposure to the target language (input) is insufficient, language learners have to absorb the material actively, processing it and using the new lexis conversationally. Language learners who both receive and produce the target language are most likely to succeed in their language learning.

Although the theories mentioned above focus on different aspects of human communication, the interactionist theories of SLA (Input, Output and Interaction hypotheses) share in the notion that meaning negotiation is critical in SLA. The Input Hypothesis argues that language learning will arise when input is comprehensible, and that input needs to extend slightly beyond the learners' current level so there is a need for learners to 'negotiate' the meaning using environmental cues. The Output hypothesis argues that language learning will succeed when learners strive to overcome gaps through negotiation of meaning. The Interaction hypothesis suggests negotiation via interaction can increase the comprehensibility and richness of any input, ensuring language learning will happen. Mitchell and Myles (2004) conclude that there are generally two types of interactionist research: one which is more interested in environmental factors associated with language learning; and the other, which investigates information processing issues (how environmental language transforms into input and then into intake). Several of the achievements of researchers investigating interactionist (Input or Interaction) hypotheses are summarised by Mitchell and Myles (2004):

1. *It has been shown that native speaker and non-native speaker interlocutors (child and adult) can and will work actively to achieve*

mutual understanding, at least when undertaking a fairly wide range of problem-solving tasks.

2. *It has been shown that these negotiations involve both linguistic and interactional modifications which together offer repeated opportunities to 'notice' aspects of target language form, whether from positive or negative evidence.*
3. *It has been shown that non-native speaker participants in 'negotiations for meaning' can attend to, take up and use language items made available to them by their native speaker interlocutors.*
4. *It has been shown that learners receiving negative feedback, relating to particular target language structures, can in some circumstances be significantly advantaged when later tested on those structures.*

(Mitchell and Myles, 2004: 191)

In this study, attention is on whether different environmental factors generate different forms of language use; that is, whether different levels of 'negotiation for meaning' in language learning result in different language performances. This study will provide additional evidence, based on research in this area.

2.2 Language learning, language use and the mind

In section 2.1 we discussed theories concerning the cognitive/psychological aspects of language learning. These express two perspectives: the first of which views language learning ability as a special aspect of mind. Researchers who hold this view frequently also support the notion that there is a 'language system' already present in our minds at birth. They try to bridge the cognitive mechanism, the processing component of language learning, with this already present mental map of language systems. They also aim to illustrate the process computationally, setting out processing approaches (Mitchell and Myles, 2004: 97).

The alternate view is that language learning ability is a single aspect of general cognitive ability, and so the way people process linguistic information is similar to how they interact with other types of information. Researchers who hold this view liken language processing to other types of information processing. Their approaches are referred to as emergentist or constructionist approaches. Unlike processing approaches, in which the processing of language learning in the mind with a computational perspective is the primary focus, constructivist/emergentist approaches draw attention to "the linguistic sign as a set of mappings between phonological forms and conceptual meanings or communicative intentions" (Ellis, 2003: 63). The constructivist/emergentist researchers "believe that the complexity of language emerges from associative learning processes being exposed to a massive and complex environment" (Ellis, 2003: 84). As Mitchell and Myles summarise:

> *Learning in this view [constructivist/emergentist approaches] is seen as the analysis of patterns in the language input, and language development is seen as resulting from the billions of associations which are made during language use,*

> *and which lead to regular patterns that might look rule-like, but in fact are merely associations.* (Mitchell and Myles, 2004: 98)

Since the current study assesses the different environmental factors leading to different language acquisition models, i.e. other than the inner processing of the mind, the following section focuses on constructivist/emergentist approaches.

Usage-Based Approach

Constructivist or emergentist views of language learning share a usage-based view of language development, as referred to above. Arguably, there is a mechanism within people's minds that supports learning, and the interaction between one's cognition and the external world (and the desire to use language) also matters. This notion, although not explicitly, echoes the interactionist SLA theories introduced in section 2.1. Therefore, it is worth discussing if we wish to investigate the relationship between language learning and immersion in/exposure to input.

The term 'usage-based' was first introduced by Langacker (1987, 1988, 2000, 2002), but has since been overly broadly used. There are two major traditions associated with usage-based theories: The Firthian tradition and enunciativist linguistics. The Firthian tradition emphasises key social aspects and context. The enunciativist linguists emphasises the speech act itself (Kemmer and Barlow, 2000: vii). Although many researchers employ the term 'usage-based approach' in their studies, the coverage of concepts in such studies is variable. To clarify, Kemmer and Barlow (2000: viii-xxii) outline nine key aspects of a usage-based model/approach:

1. *The intimate relation between linguistic structures and instances of use of language.*
2. *The importance of frequency.*
3. *Comprehension and production as integral, rather than peripheral, to the linguistic system.*
4. *Focus on the role of learning and experience in language acquisition.*
5. *Linguistic representations as emergent, rather than stored as fixed entities.*
6. *Importance of usage data in theory construction and description.*
7. *The intimate relation between usage, synchronic variation, and diachronic change.*
8. *The interconnectedness of the linguistic system with non-linguistic cognitive systems.*
9. *The crucial role of context in the operation of the linguistic system.*

(Kemmer and Barlow, 2000: viii-xxii)

The following section discusses each of these nine key aspects in turns.

1. The intimate relationship between linguistic structures and instances of use of language.

Clearly, since the principal idea informing this approach deems the interaction between one's cognition and the world at large as central, a researcher choosing to apply a usage-based approach would be expected to conduct an investigation of linguistic structures by investigating instances of language use. That is, a speaker's linguistic system is essentially based on 'usage events', which can be defined as "the pairing of a vocalisation, in all its specificity, with a conceptualisation representing its full contextual understanding" (Langacker, 2000: 9). There are three points to make about the relationship between one's linguistic system and usage events. Firstly, a speaker's linguistic system is formulated

according to instances obtained in ‘usage events’. Secondly, the abstract form of a speaker’s grammar, and the instances experienced by the speaker previously are intertwined. Thirdly, the influence of ‘usage events’ is ongoing in linguistic systems in a broad sense; that is, one’s production of language influence not only relates to the products of one’s linguistic system, but also influences other speakers’ linguistic systems. This type of influence does not just affect one’s initial acquisition of language, but is lifelong (Kemmer and Barlow, 2000: ix).

2. The importance of frequency.

Given that the usage-based approach assumes language acquisition is an experience-driven/usage event-driven, the frequency of instances of experience play an important role in any research. As the theory of memory (information processing theory) suggests, people have three types of memory storage: (1) sensory storage, which contains a trace of outer stimuli (sounds/words) linked to inner cognition (meanings/patterns); (2) short term memory (working memory), in which instant and temporary information is processed; and (3) long term memory, in which knowledge is stored (Atkinson and Shiffrin, 1968). When one pattern of stimuli and meanings/information occurs frequently, it is likely to become stored in a person’s long-term memory. Therefore, the frequency of instances of language experience are seen to be important within this cognitive approach to language acquisition.

3. Comprehension and production as integral, rather than peripheral, to the linguistic system.

In a usage-based approach, a speaker’s linguistic system is formed and operated according to usage events; therefore, “the structure of this [linguistic] system is not separate in any significant way from the (cumulative) acts of mental processing that occur in language use” (Kemmer and Barlow, 2000: xi). In other words, here the traditional separation of ‘competence’ and ‘performance’ is rejected, and a speaker’s language performance is judged

equivalent to his/her competence in the mental process sense.

4. *Focus on the role of learning and experience in language acquisition.*

Cognitive linguists acknowledge the importance of learning mechanisms. Their usage-based approach deems a speaker's experience to be of equal importance to learning mechanisms, since the linguistic system itself is influenced by instances of language production (output) and comprehension (input).

5. *Linguistic representations as emergent, rather than stored as fixed entities.*

The development of a usage-based approach arose from the work of cognitive linguistics. The general linguistic view suggesting a linguistic system is a storage of a set of fixed units utilised according to a set of stored and fixed instructions or rules is not accepted by constructivist/emergentist researchers:

> *Linguistic units are seen as cognitive routines. Such units are nothing more than recurrent patterns of mental (ultimately neural) activation […] During linguistic processing, linguistic units are part and parcel of the system's processing activity: they exist as activation patterns. When no processing is occurring, the information represented by such units simply resides in patterns of connectivity (including differential connection strengths) resulting from previous activations.* (Kemmer and Barlow, 2000: xii-xiii)

6. *Importance of usage data in theory construction and description.*

According to the usage-based approach, the linguistic system is constructed in response to frequent exposure to 'usage events', and authentic language is seen as a key element in acquisition. Thus, the authentic language that people produce (output) and comprehend (input)

influences the way they use language.

7. The intimate relation between usage, synchronic variation, and diachronic change.

"The selection of a given entrenched variant for activation is governed by a complex set of motivating factors, including system-internal as well as contextual, situational factors" (Kemmer and Barlow, 2000: xviii). Thus, as long as a 'usage event' shifts in response to time and space, contextual factors or linguistic system-internal factors will also adapt accordingly. These changes do not occur only within an individual system, but also influence a group of people who share the same community. As Kemmer and Israel (1994: 167) state, "the more speakers talk to each other the more they will talk alike, and so linguistic variation will pattern along lines of social contact and interaction." (The existence of dialects evidences this claim.)

8. The interconnectedness of the linguistic system with non-linguistic cognitive systems.

In this view of cognitive theory, language is conceived as a sub set of cognitive systems, connected to other cognitive concepts in use. For instance, in Fillmore's work on frame semantics, he observes that the situation of language use often incorporates additional linguistic information that conventional linguistic units (such as words and grammatical constructions) cannot fully cover (Fillmore, 1977; Fillmore and Atkins, 1992).

9. The crucial role of context in the operation of the linguistic system.

The usage-based approach suggests "it is highly likely that both linguistic and non-linguistic patterns will be processed and learned in an integrated way" (Kemmer and Barlow, 2000: xxi). Thus, when people encounter a situation involving language use, all aspects of language (from phonetics to semantics) are then influenced by both linguistic and non-linguistic context. The linguistic input taken from context can influence people's

language use, e.g. their pronunciation of words. In addition, non-linguistic elements from context can also influence people's language use, e.g. the social status of interlocutors, or the formality of the conversation.

To summarise the previous statements, we show that in a usage-based approach, authentic instances of language use are important to the process of language acquisition, since one's linguistic system is constructed based upon it. Therefore, it is also essential to use authentic language data when studying language acquisition. The frequency of instances of language use serve as a crucial factor enabling people to organise and interpret instances of language use. What a speaker has in his/her linguistic system is also what s/he presents in terms of language use and vice versa. Since the construction of linguistic units is a reaction to an immediate 'usage-event' a speaker encounters, usage-based theorists argue no fixed lexical units exist to accompany fixed grammatical rules in a conventional linguistic sense. People's language use proceeds from patterns of language use speakers generate based on their previous experience of stimuli and meanings from various contexts. Changes in language use happen both synchronically and diachronically. The use of language is part of a larger set of cognitive concepts, as there are always other cognitive concepts involved that function beyond the domain of language. All aspects of a usage-based approach stress context as significant in studies of language acquisition.

The usage-based approach, as described herein, can be seen to complement the SLA theories discussed in 2.1 and vice versa. The Input Hypothesis emphasises the importance of comprehensible input in SLA, and the Interaction Hypothesis points out that comprehensible input arises from 'negotiation' between language users. The Output Hypothesis further argues that comprehensible input (receiving) cannot help learners succeed in isolation, it is rather the use of language (producing) after input is received that reinforces learners' newly learnt

knowledge of language. The key concepts in a usage-based approach, authenticity and frequency of language instances (input), can be seen as a supplement to the concept of comprehensible input and the reinforcement of language learning through output production. Engagement of language use events in person provide outer world/linguistic stimuli and inner world/linguistic knowledge to support one's information processing as regards language. From the perspective of information processing theory, a pattern-forming mechanism arises in the mind. When a pattern is encountered frequently, it is likely to become stored in the memory for easy retrieval subsequently. Swain's Output theory might be explainable in terms of this idea. When a learner first accesses a language pattern, there is only a single trace of this stimulus. If, however, the learner then has to produce the language pattern themselves, this trace of stimulus then needs to be retrieved and analysed by the mind several times; the result is that the pattern is then likely to be stored more efficiently.

The usage-based approach also provides a way of seeing how input from the environment/context and the output of language can best be investigated. The concepts relevant to language use development are: (1) when people encounter a situation of language use, all aspects of language (from phonetics to semantics) can be influenced by both linguistic and non-linguistic context causing changes in language use to occur; and (2) there are no fixed lexical units necessary to accompany fixed grammatical rules in the conventional linguistic sense. People's language use is based on the patterns they generate in response to their previous experience of stimuli and meanings in various contexts. In other words, analysis of language use should focus on people's language use patterns (small units of language) and take contextual factors into consideration.

The view when analysing language in terms of patterns/smaller units can be linked to studies of lexico-grammar, lexical phrasing or prefabricated language (Nattinger and

DeCarrico, 1989, 1992) and the Lexical Approach (Lewis, 1993). In Nattinger and DeCarrico's definition, lexical phrases are 'chunks of language' that vary in length:

> *[Lexical phrases are] multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax, conventionalised form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time.* (Nattinger and DeCarrico, 1992: 1)

Nattinger and DeCarrico (1992) suggest "ritualisation" is crucial in regard to language, and prefabricated chunks of language are the production of "ritualisation". As Lewis (1993: vi) states, "language consists of grammaticalised lexis, not lexicalised grammar." Nattinger, DeCarrico and Lewis all opine that language is 'built up' of multi-word units (linked to phraseology, the study of the collocation of words) that are joined together ritually according to frequency and idiomaticity.

Collocation is defined as "an aspect of lexical cohesion which embraces a 'relationship' between lexical items that regularly co-occur" by Carter (1988: 163). Sinclair (1991: 115) states of collocation that "words appear to be chosen in pairs or groups and these are not necessarily adjacent." Meanwhile, Hunston (2002: 12) defines collocation as "the statistical tendency of words to co-occur", and Nunan (2012) describes a key characteristic of collocation as words that "frequently co-occur". Herein we consider collocation to refer to 'words which show a statistical tendency to co-occur'. This tendency is often expressed by word frequency and analysis of frequency is a crucial aspect of the study of phraseology. As Teubert (2005) states:

> *Frequency is an important parameter for detecting recurrent patterns defined by the co-occurrence of words. Frequency is thus an essential feature for making general claims about the discourse.* (Teubert, 2005: 5)

Multi-word units combine not merely in terms of the frequency of their appearance together, but also due to idiomaticity. Discussing idiomaticity, Sinclair (1991) mentions two principles for interpreting the concept in texts: one is the open-choice principle and the other is the idiom principle.

The Open-choice Principle and the Idiom Principle

Sinclair (1991) suggests language is actually "the result of a very large number of complex choices." The open-choice principle is also referred to as a 'slot-and-filler' model. It is believed there are series of slots in texts, and these slots need to be filled with words from the lexicon. However, this process of choosing words is not random. Sinclair suggests that global knowledge and environmental factors influence how words are selected. Words that co-occur frequently in common understanding may have a strong probability of being mentioned together, e.g. *criminals* and *police*, or *rose* and *red*. Similar or opposing concepts are also often mentioned together frequently, such as *love* and *peace*, or *anxiety* and *relief*. Under certain social conditions and contexts, some words are more likely to be collocated. For example, in campaign literature, words such as *trust, believe* or *promise* are commonplace (Sinclair, 1991: 109f.).

In terms of the idiom principle, Sinclair describes it as based on "a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to

be analysable into segments" (1991: 110). Hunston offers further explanation of this principle:

> *…that is, each word in the text is used in a common phraseology, meaning is attached to the whole phrase rather than to the individual parts of it, and the hearer or reader understands the phrase as a phrase rather than as a grammatical template with lexical items in it.* (Hunston, 2002: 143)

In contrast to the open-choice principle, this indicates that when a speaker chooses an item from the lexicon, word choice is not limited by the grammaticality alone, but by a more complex mechanism. These semi-preconstructed phrases (multi-word units) have unique restrictions that they are bound to. The multi-word units are usually chosen in a conventional way; however, unlike traditional idioms, which are totally fixed and scarcely vary in form (e.g. *kick the bucket*), these semi-preconstructed phrases (multi-word units) are, to some extent, allowed to incorporate different means of expression (Sinclair, 1991), e.g. *to be frank/honest* or *what [I/you] [say/mean] is that…*.

Formulaic language theory

In a study of the idiomaticity of multi-word units, Wray (2002) states that adult native speakers do not generate or analyse every piece of their language with full comprehension every time. Idioms are frequently difficult to analyse based on their literal meanings. As Pawley and Syder (1983) point out, although theoretically people can use all possible grammatical sentences, they are more likely to use certain types of grammatical sentence than others. Wray (2002) illustrates this with "*The captain has illuminated the seatbelt sign as an indication that landing is imminent*". She argues that even though native speakers might use this grammatical sentence, the expression "*The captain has put the seatbelt sign on, which*

means we're about to land" is a more likely and natural format (Wray, 2002: 13). Thus, phrases/sentences are not constructed freely. As Sinclair (1991: 110) states, there is an "unrandomness" in the distribution of words, that suggests people process language according to both the open choice principle and the idiom principle. Similarly, Wray (1992, 2002) suggests that according to *formulaic language theory*, there are two systems involved in language use: the analytic processing system and the holistic processing system. Wray illustrates these as follows:

> *Analytic processing entails the interaction of words and morphemes with grammatical rules, to create, and decode, novel, or potentially novel, linguistic material. Holistic processing relies on prefabricated strings stored in memory.* (Wray, 2002: 14)

Wray further suggests people use these two processing methods according to discourse context. Moreover, the prefabricated word strings in holistic processing are not completely fixed either:

> *…holistic processing is not restricted to only those strings which cannot be created or understood by rules, such as idioms. It can also address linguistic material, for which grammatical processing would have rendered exactly the same result.* (Wray, 2002: 14)

Reviewing the theories proposed by Sinclair (1991) and Wray (2002), it is apparent that both place prefabricated language at the centre of studies on language use. Sinclair (1991) argues that there is no clear-cut distinction between patterns and meanings, or between lexis and grammar. Nattinger and DeCarrico (1992) echo this notion below:

> *[T]here is an entire range of items, some of which are specific and pertain to a small number of instances (lexical items), and some of which are very general and pertain to a large number of instances (grammar rules). But since elements exist at all levels of generality, it seems impossible to draw a sharp border between them.* (Nattinger & DeCarrico, 1992: 22)

Nattinger and DeCarrico (1992) use the term *lexical phrases* to identify the "lexico-grammatical units that occupy a position somewhere between the traditional poles of lexicon and syntax" (Nattinger & DeCarrico, 1992: 36). They observe that *lexical phrases* are form/function composites that differ from idioms. *Lexical phrases* are used for certain functions, and idioms seldom have this kind of function. For instance, the idiom *kick the bucket* only means someone has died, and the lexical phrase *the* ____*er the* _____*er* is used to express comparative relationships in discourse. Nattinger and DeCarrico (2002: 36) point out that *lexical phrases* need to be viewed separately from the general collocations of words. In their opinion, collocations are linear units of particular lexical items that co-occur frequently (not randomly), and usually contain little or no pragmatic function. For instance, the collocation *black and white* is more frequent than *white and black*, although it contains no pragmatic functions. Nattinger and DeCarrico (2002: 36) further state that *lexical phrases* are collocations with pragmatic functions, identifying two major types:

(1) *Strings of particular lexical items/non-productive lexical items*:

The lexical items in strings cannot be substituted with other paradigmatic or syntagmatic lexical items. Some of these strings conform to a grammatical/syntactic string, e.g. *at any rate*; and some do not conform to a grammatical/syntactic string, e.g. *by and large*.

(2) *Generalised frames/productive frames*:

Strings consist of specified syntactic/semantic sets of lexical items, or particular lexical items, comprising pragmatic functions. For instance, *a + N* [*+ time*] *+ ago* (*a decade ago*) and *Modal + you + VP* (*could you open the door?*) are two canonical examples of generalised frames. They can also be non-canonical, e.g. *Adv.* [*+ direction*] *+ with + NP* (*off with his head*).

Similar ideas for strings of particular lexical items (with or without pragmatic functions/generalisable or stand-alone) lie between the continuum of 'openness' to 'idioms' and are discussed by other researchers in different terms, e.g. *chunks, collocations* (Sinclair, 1991), *fixed expressions* (Hudson, 1998), *lexical phrases* (Nattinger and DeCarrico (2002) and *formulaic sequences* (Wray, 2002; Schmitt, 2004). Wray (2002) and Schmitt (2004) use the term *formulaic sequences* when discussing the formulaicity of language. Wray's (2002) definition of a *formulaic sequence* is:

> *…a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.* (Wray, 2002: 9)

Schmitt and Carter (2004: 4) state that the term *formulaic sequence* covers a wide range of phraseology, and is linked to the following characteristics (Schmitt and Carter, 2004: 4-10):

1. *Formulaic sequences appear to be stored in the mind as holistic units, but they may*

not be acquired in an all-or-nothing manner.

2. *Formulaic sequences can have slots to enable flexibility of use, but the slots typically have semantic constraints.*
3. *Formulaic sequences can have semantic prosody.*
4. *Formulaic sequences are often tied to particular conditions of use.*

Schmitt and Carter (2004) draw their first conclusions based on evidence which suggests that "formulaic sequences are typically stored and processed as unitary wholes" (2004: 4), e.g. idioms or short phrases such as *rain cats and dogs* or *go away*. They point out that a fixed formulaic sequence will be learnt first, and later analysed to form a formulaic sequence with slots. Moreover, short formulaic sequences are likened to raw materials, which can be blended to create longer and more complex forms.

The second conclusion is based on the observation of formulaic sequences in which slots are used within a fixed frame, e.g. *a* _____ *ago* (*a decade ago/a year ago*). The slots in this type of formulaic sequence are often perceived as semantic constraints. For instance, the formulaic sequence *a* ____ *ago* requires a noun, which is then relevant to time. In the third conclusion, it is noted that certain words can be perceived with positive or negative associations based on the frequent occurrences of specific collocations. For instance, the verb *cause* is most often used in a negative context (e.g. *cause disaster/accident*). This feature is termed 'semantic prosody' (Sinclair, 2004) and establishes that the slots used in formulaic sequences have semantic constraints, often conveying semantic prosody. The fourth conclusion proceeds from the observation that formulaic sequences are useful for many functional purposes. For example, formulaic sequences can be used to maintain social interactions, such as agreeing (*Okay, I see*), or for discourse organisation, e.g. re-phrasing *as I*

was saying. Therefore, formulaic sequences are often associated with certain conditions of use (Schmitt and Carter, 2004: 4-10).

The acquisition of formulaic sequences

When viewing language acquisition from the perspective of formulaic language, it emerged that children who prefer to focus on things/activities are likely to learn more single words (more nouns), and children who prefer to interact socially learn more expressions in whole (more non-segmented phrases) (Nelson, 1973). This finding links the desire to interact to the use of formulaic sequences. This first language acquisition behaviour reflects that found in L2 acquirers, who aim to learn formulaic sequences to meet their need to communicate (Schmitt and Carter, 2004: 11-12). Wood (2002: 5 cited in Schmitt and Carter, 2004: 12) summarises the role of formulaic sequencing in language acquisition:

> *[Formulaic sequences] are acquired and retained in and of themselves, linked to pragmatic competence and expanded as this aspect of communicative ability and awareness develops. At the same time, they are segmented and analysed, broken down, and combined as cognitive skills of analysis and synthesis grow. Both the original formulas and the pieces and rules that come from analysis are retained.* (Wood, 2002: 5)

Undoubtedly, intention to communicate socially is a key motive governing language acquisition. In terms of the process of acquisition, Wray and Perkins (2000) and Wray (2002) argue that language acquisition involves sequence-based learning and grammatically-generated learning. Moreover, sequence-based language and grammatically-generated language vary in terms of distribution

during an L1 child's development (Schmitt and Carter, 2004: 12). Formulaic sequence development starts after a large amount of vocabulary (single words or phrases) has been memorised, as the use of analytic language starts to increase.

> *Language production increasingly becomes a top-down process of formula blending as opposed to a bottom-up process of combining single lexical items in accordance with the specification of the grammar.* (Wray and Perkins, 2000: 21)

The development of formulaic sequencing in SLA differs from that experienced by L1 learners. In terms of idiomatic language, L2 learners often use it in a very limited way, or even avoid using it completely. It is possible that a lack of idiomatic language input might explain this phenomenon (Kellerman, 1978; Irujo, 1986). It has also been found that L2 learners overuse, underuse, or misuse some formulaic sequences when compared to L1 speakers' use of formulaic sequences (De Cock, 2000). It is difficult to generalise a clear procedure for L2 acquisition in formulaic language due to the diversity of L2 learners. Differences in learners' ages, first language, learning methods, environment and other factors can affect how L2 learners acquire formulaic sequences (Wray, 2002).

According to the scale of information discussed in terms of formulaic language (lexical phrases/formulaic sequences), it is difficult to locate a clear border between syntax and lexis. As our previous discussion indicates, restrictions on word choice vary according to different kinds of multi-word units, and so language acquisition seems to be a 'top-down' process. This poses the question of whether language acquisition really starts with learning grammatical rules, as perhaps there is another angle to view how language is used by people. Ellis (1996, 2002) suggests people learn a new language according to their ability to deduce patterns from

language input. This implies language acquisition is pattern-based, and not determined by the innate principles and parameters of an inborn linguistic system. People learn that combinations of phonemes/morphemes/words are acceptable in the target language, and consequently become aware of what combinations of phonemes/morphemes/words appear to be 'natural' or 'unnatural'. Explicating pattern-based language acquisition, Hoey (2005) presents a theory called *lexical priming theory*.

Lexical priming theory

As mentioned above, lexical priming theory was proposed by Hoey (2004a, 2004b, 2005, 2008). Similar to cognitive linguists and other researchers studying phraseology/formulaic sequences, he argued that the traditional view of language, influenced greatly by Chomsky's view of grammar since the 1960s, fails to explain the requirements for naturalness in language, i.e. why grammatically correct sentences can seem unnatural, proposing collocation as the key to understanding naturalness. To exemplify this, Hoey (2005) offers two versions of a sentence from a book by Bill Bryson, one is Bryson's original, and the other a grammatically acceptable variant, to demonstrates that the fabricate sentence seems clumsy because the collocations used are low frequency:

> [1] *In winter Hammerfest is a thirty-hour ride by bus from Oslo, though why anyone would want to go there in winter is a question worth considering.* [original sentence in Bill Bryson (1991); quoted in Hoey (2005: 5)]
>
> [2] *Through winter, rides between Oslo and Hammerfest use thirty hours up in a bus, though why travellers would select to ride there then might be pondered.* [fabricated sentence by Hoey (2005: 5)]

Hoey's theory draws heavily on corpus linguistics, which concentrate on three key elements, as follows (Hoey, 2005; Pace-Sigge, 2013: 11):

1. *Collocation*: words that are found in close proximity to each other, like *naked eye*, or *you know*.
2. *Colligation*: words that form a direct grammatical construction, e.g. *that winter* is always used for past tense constructions, whereas *in the winter* is only used in the past tense half the time.
3. *Semantic association*: *that winter* refers to a 'specific event', while *in winter* mostly refers to a 'timeless truth'.

Collocation, according to Hoey's definition, is "a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution" (Hoey, 2005: 5). The term colligation is defined as

> *…the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank; the grammatical functions preferred or avoided by the group in which the word or word sequence participates; the place in a sequence that a word or word sequence prefers (or avoids).* (Hoey, 2005: 43)

Based on these definitions, it becomes apparent that Hoey's theory draws on both corpus linguistics and psychological findings. From investigations of corpus texts Hoey concludes that "lexis is complexly and systematically structured and that grammar is an outcome of this

lexical structure" (Hoey, 2005: 1); therefore, it is more reasonable to posit lexis as at the core of language theory.

The study of collocation is often associated with phraseology, which is linked to the growing popularity of corpus linguistics, which is an important feature of corpus linguistic research. As Guy Cook states, corpus linguistics concerns "patterns and regularities of language use which can be revealed by systematic analysis of corpora. One of the most important insights relates to collocation" (2003: 73). Hoey did not devise lexical priming theory from phraseology and corpus linguistics alone; he also commented on 'the pervasiveness of collocation', noting that the interlocking collocations observed in some sentences suggest a lexicon-based theory as a basis from which to investigate whether collocation indicates more than simply the co-occurrence of two or more words. The psychological features of collocation in Hoey's definition are extended when the concept of priming in psychology is adapted and modified to explain collocation.

The term *priming* in psychology also refers to a "factor that influences the accessibility of information in memory" (Sherman *et al.*, 2003: 55). This can be explained thus,

> *…the activation of stored knowledge through experiences in the immediate context can make prime-relevant information more accessible in memory, and such recent construct activation can influence inferences, evaluations, and decisions on subsequent tasks.* (Sherman *et al.*, 2003: 55)

To illustrate; if one is shown several photographs of dogs and then asked to write down a word with the letter D in the initial position, it is very likely one will write down the word *dogs*. Early psychological studies of priming examined semantic information only. Collins and Quillian (1969, 1970) conducted a series of experiments showing that people reacted more quickly when identifying 'true sentences' (*tennis is a game*), than 'false sentences' (*football is a lottery*). These findings suggest concepts and their meanings in semantic memory are activated through the mechanism of priming. Hoey takes the concept of semantic memory assuming that:

> *…every word [and word sequence] is mentally primed for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context.* (Hoey, 2005: 8)

Although Hoey does not explicitly state that lexical priming theory relates to language acquisition, its relevance is apparent. Indeed, he does discuss children's acquisition of priming:

> *…a child acquires the primings of a combination first and the primings of the individual words latter (e.g. all gone). There is no difference in principle between acquiring the word (or word sequence) and acquiring the knowledge of its collocation.* (Hoey, 2005: 8)

Hoey suggests lexical priming theory bridges between the language in a person's brain and the language a person utters externally, which can be deemed a cognitive linguistic claim. He also confirms that the notion of priming can be applied to Giddens' (1979) view concerning individuals and social structure, which states that every individual action recreates social structures, and those social structures in turn shape individual action (Hoey, 2005: 8). Therefore, this theory might be seen as a social-cognitive relevant theory of language acquisition. Moreover, Hoey (2005: 9) states that his view of grammar shares features with Hopper's 'emergent grammar' (1988, 1998), which claims that grammar results from 'routine'; i.e. the repeated encounter and use of identical language elements. When patterns/grammar get formulated they can then differ from individual to individual. This concept of emergent grammar is similar to the fifth description of the usage-based approach and Ellis's (1996, 2002) pattern-based models of language acquisition, as discussed in the previous section.

Vocabulary knowledge within the skills of writing, reading, listening and speaking is common to be divided into two kinds: productive and receptive knowledge. (Nation, 2001; Webb, 2005, 2008). Receptive vocabulary knowledge works when learners read or listen to texts. Learners know and recognise the meaning of words which leads them to understand the texts. However, this knowledge of vocabulary may not be used within the skills of speaking and writing. It is the productiove vocabulary knowledge that works when learners speak or write texts. Pruoductive vocabulary knowledge is the abiltiy to retrieve the knowledge of the structure and meaning of words (Webb, 2005, 2008). Lexical priming theory, similarly, regards that there are two kinds of primings involved. One is productive primings, and the other is receptive primings. When a learner encounters a word or word sequence repeatedly in discourse and takes part in producing this word or word sequence, s/he then develops

productive primings of the word or word sequence. If a learner encounters a word or word sequence in context where getting involved actively seems impossible, such as watching a TV show or reading a novel, s/he then develops receptive primings of the word or word sequence. (Hoey, 2005: 11)

Lexical priming theory is also a theory of language use, which explains language use is constantly influenced by external contexts. "Priming need not be a permanent feature of the word or word sequence" (Hoey, 2005: 9). That is, people's primings of a word (or word sequence) are never fixed, as every time a person encounters a word/word sequence, the associated contexts are new. Even when they are similar to other contexts, they are not identical. Therefore, primings can be weakened or reinforced according to the experience of a person who has encountered the word/word sequence repeatedly. That is, a drift in the priming can occur. This is the mechanism for language change; i.e. change occurs when a sufficient number of members of a community experience a drift in their primings simultaneously (Hoey, 2005: 9). Lexical priming theory was further developed to cover a wide range of linguistic features, as shown in Hoey's hypotheses below.

Priming hypotheses

Every word is primed for use in discourse in response to the cumulative effects of an individual's encounters with the word. If one of the effects of initial priming is that regular word sequences are constructed, then these are also in turn primed. More specifically:

1. Every word is primed to occur with particular other words; these are its collocates.
2. Every word is primed to occur with particular semantic sets; these are its semantic associations.
3. Every word is primed to occur in association with particular pragmatic functions;

these are its pragmatic associations.

4. Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.
5. Co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations.
6. When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.
7. Every word is primed for use in one or more grammatical roles; these are its grammatical categories.
8. Every word is primed to participate in, or avoid, particular type of cohesive relation in a discourse; these are its textual collocations.
9. Every word is primed to occur in particular semantic relations in discourse; these are its textual semantic associations.
10. Every word is primed to occur in, or avoid, certain positions in discourse; these are its textual colligations.

Crucially, all these claims are in the first instance constrained by domain and/or genre.

(Adapted from Hoey, 2005: 13)

In summary, lexical priming theory is a language theory that (1) assigns to lexis and establishes the collocation of lexical items as pivotal in language, (2) is concerned to explain actual use and naturalness of language, instead of describing merely the abstract possibility of how language may be used, (3) borrows and applies the psychological concept of priming to conceptualise and describe the property of language in a flexible and extendable manner, which can then potentially be applied at all linguistic levels (morphological, semantic, grammatical, discourse/textual, and pragmatic), (4) studies authentic language data in large quantities, using corpus linguistic methods, and 5) provides a potential explanation for language acquisition and use.

In light of our previous discussion regarding usage-based approach, we can see that lexical priming theory, although it does not focus on the cognitive study of language, shares some concepts. Firstly, both theories value authentic instances of language and suggest one's language knowledge is constructed on the basis of authentic language exposure. Secondly, they deem the frequency of language input as crucial to language acquisition and production. Thirdly, as lexical priming theory can potentially bridge one's language in mind and language that is uttered externally, it is close to the notion of a usage-based approach which states that the knowledge a speaker has in his/her linguist system is apparent from how s/he presents it in language use. Fourthly, usage-based theorists argue there are no fixed lexical units available to accompany fixed grammatical rules in a conventional linguistic sense, and nor are the primings of words/word sequences permanently fixed. Fifthly, both suggest that changes in language use can happen synchronically and diachronically, as Hoey states, "priming is the driving force behind language use, language structure and language change" (2005: 12). Sixthly, both the usage-based approach and lexical priming theory suggest the context in which language input arises plays an important role in language acquisition.

The comparison above, of course, does not imply that lexical priming theory is another usage-based theory of language, since lexical priming theory proceeds from linguistic studies of phraseology and corpus linguistics, adding a twist in the form of a psychological concept (priming). The similarity between a usage-based approach and lexical priming theory may result from the fact that Hoey's theory shares many commonalties with Sinclair (1991, 1997, 2000, 2004) and Stubbs (1995, 1996, 2001). Also, the idea of priming in psychology highlights the importance of context.

As Kemmer and Barlow state (2000: xxii):

> *The importance of context and in particular the social aspects of context for understanding the form and nature of language has historically been more of a major feature of British and other European linguistic traditions […] In Firthian linguistics in particular, as mentioned earlier, context plays a key role. This tradition has been continued in work by linguists such as John Sinclair and Michael Stubbs (e.g. Sinclair 1991, Stubbs 1996), who not only examine textural patterns such as collocations, but also the context of use of such patterns whether relating to register, institutions, or culture.* (Kemmer and Barlow, 2000: xxii)

Despite lexical priming theory making some claims about language acquisition, it is rarely linked to language acquisition/SLA explicitly. After all, lexical priming theory initially developed as a way to explain language use, as observed in corpus data. Through comparison with cognitive linguistic theory, and showing the similarities shared by both, it reveals the extendable and flexible characteristics of lexical priming theory. However, whether this theory can genuinely serve as a theory of language acquisition warrants further investigation.

As has been discussed in relation to the acquisition of second language, language input, output, interactions between interlocutors, and other factors (such as the diversity of ages, first language, learning methods, environment) all inform how people acquire and use language. Krashen's (1985) affective filter hypothesis also acknowledges that the acquirer's attitude/emotion can also affect language acquisition. Priming involves multiple factors, all of which influence ones' access to memorised information which was laid down and experienced in various contexts. These contexts comprise numerous environmental situations, social events, and communicative purposes. One's physical/mental condition also plays a role in receiving, processing and storing memories. All these factors affect how a person is primed

to access information via their memory. Therefore, theoretically, such factors also influence how a person is primed to use particular words/word sequences.

The fundamental issue that arises here is that solid evidence is required to establish the existence of lexical priming. Lexical priming theory borrows the concept of 'priming' from psychology, where it is used to describe people that are prone to group frequent/similar instances together and prone not to group infrequent/dissimilar instances together. This theory aims to bridge the propensity for particular language use observed in data sets, to explain what might be happening in language users' minds. It is hoped that describing and explaining people's language use from a psychological perspective, can be supported by empirical psychological studies, which provide potential evidence for the veracity of lexical priming theory. For example, Collins and Quillian (1969, 1970, 1972) link semantic memory to the process of priming, as outlined at the start of this section in relation to true and false statements. Elsewhere, Meyer and Schvanefeldt (1971) conducted an experiment requiring participants to classify words as 'unassociated words' or 'related words'. They found people react faster to pairs of associated words than to pairs of unassociated words. This finding shows "degree of association is a powerful factor affecting lexical decisions" (Meyer and Schvanefeldt, 1971: 229). These findings can, to some degree, support the claim that "Every word is primed to occur with particular other words/with particular semantic sets" (Hoey, 2005: 13) in lexical priming theory. An observation of brain-damaged patients was made by De Mornay Davies (1998), who worked with patients who struggled to access the semantic information associated with a target word; i.e. patients who could not automatically retrieve information from their semantic memories. De Mornay Davies (1998) used the term "Automatic semantic priming" to describe the process involved, highlighting its relevance to the lexicon held in the memory:

> *Automatic semantic priming assumes that, on presentation of a word, the information about that word is retrieved as a result of lexical access, rather than being retrieved explicitly as a result of subjects' responses to task demands.* (De Mornay Davies, 1998: 391)

The findings of De Mornay Davies (1998: 402) led to the assertion that there is a 'lexical and semantic automatism', that it is activated "by co-occurrence frequency, such that words which often co-occur in speech or text." This echoes the description of priming in lexical priming theory and asserts that priming is an automatic process within a person's language system. However, the linguistic elements tested in these studies were mostly single words and their semantic associations, and the methods implemented used psychological experiments.

More pertinently, Gries (2005) and Ellis et all. (2008, 2009) applied corpus-based methods to evaluate people's priming and expand the scope of linguistic elements implicated. The research conducted by Gries (2005) analysed two pairs of syntactic patterns in the ICE-GB corpus to ascertain whether corpus data matched the results of psychological experiments. He found:

> *[T]he corpus-based analysis of syntactic priming revealed significant priming effects for ditransitives and prepositional datives, the results are also strikingly similar to those of previous experimental studies in terms of strength of effects, the influence of morphological characteristics of the verb, construction-specificity, directionality and distance effects (i.e. the time course of priming).* (Gries, 2005: 373)

In their research, Ellis et al. (2008) observed language produced by native/non-native ESL speakers, and pointed out that patterns which occurred with the greatest frequency were processed with high fluency. Non-native ESL speakers recognised those word strings to which they have been frequently exposed. In contrast, native speakers did not merely recognise them, they were also willing to process flexible forms of word strings. Ellis et al. (2009) investigated the frequency and strength of collocations in BNC, and applied a lexical decision paradigm to identify any processing effects. To achieve this, frequently occurring collocates/word clusters (e.g. frequent: *lose weight*; infrequent: *receive virginity*) were extracted from the corpus, and the researchers then measured people's reaction time as they made a lexical decision. It was found that frequently occurring collocations resulted in faster language processing. This demonstrated that "priming is not down to something based on semantic generalisations but is more likely due to automatic decisions made because of word associations in the memory" (Pace-Sigge, 2013: 53). Consequently, on the basis of this evidence it is reasonable to suggest that corpus studies do, to some extent, confirm the findings obtained from psychological studies. Similarly, the nature of language comprehension and language production, as apparent from corpus linguistic studies can also be confirmed in relation to psychological findings (Pace-Sigge, 2013: 53). Thus, the findings in the current study potentially support discussions about psychological priming as regards linguistic elements.

2.3 The study hypotheses

Following the summary of relevant research and theories in this chapter, this section lists the hypotheses that will be investigated in this research. In the introduction, I raised the issue of fossilisation and the SLA assumptions concerning the benefits of immersion in the target language. This chapter has discussed the role of input in SLA and how input from the external environment influences people's language acquisition. Additional environmental factors (the adjustment/negotiation from output and interaction) were also observed to influence people's language acquisition. In this study, we proceed from the belief that people's acquisition of language starts when collecting/mimicking lexical items/multi-word units obtained from communitive/comprehensible input, as people use their inborn cognitive ability, and pattern-forming ability, to generate rules of patterns/frames/formulaic sequences from examples of frequently-used language. To evaluate whether the environment and input genuinely influence learners' language use, an investigation into EFL learners different input backgrounds will be conducted to yield useful and novel findings. To observe what aspects of people's language use are influenced by input, the focus of our analysis will be on lexical units and their collocations.

My hypotheses as set out in this study are as follows:

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant

subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

Chapter 3 Methodology

This chapter describes the materials and methods used in this study. Section 3.1 briefly illustrates the research aims and explains how the research methods were designed to achieve them. Section 3.2 explains how the corpus used for research was built and describes the study participants. Section 3.3 details the method for compiling the data and illustrates the tools used for transcribing the conversation materials into analysable (see 3.3.1). The details of the corpus are also presented in this section, and 3.3.2 identifies the reference corpora used in this research and justifies their selection. Section 3.4 introduces the method of analysis (see 3.4.2), and specifies the words analysed in the study and the reasons for choosing them (see 3.4.1).

3.1 Overview

In discussion presented in the previous chapter, it emerged that when people acquire language, input is a key factor supporting learning. Interaction between interlocutors is also vital. It was also stated that native speakers typically use certain words together and generalise similar utterances according to patterns/formulaic language. This language behaviour is based on recognition from experience that some words or combinations of words co-occur more frequently or less frequently in specific contexts. Thus, Chapter 2 revealed that of the multiple factors that might influence acquisition, immersion in the target language environment and the diversity of school education are thought to be especially important.

The first factor, the immersion environment, when broadly defined, could be as limited as working in a new company where people use the target language, or as major as a move to another country where people use the foreign language. In the former scenario, although everyone uses the language within the company, the influence on the language user might be less significant. However, the non-native speaker could still acquire new jargon from colleagues. In this scenario, there are various possibilities. For example, for someone who might have learnt the foreign language in his/her own country and then moved to another country, four possibilities might arise: (i) s/he uses the foreign language s/he was taught in his/her country in the new country without any resultant influence on language use (A→A; (ii) s/he uses the language as previously taught, but integrates some aspects of the language spoken in the new setting (A→A-1); (iii) s/he uses what s/he was taught, but adopts language from the new setting (A→A+1); (iv) or, s/he completely abandons what s/he was taught previously and adopts new phrasings and words, altering their foreign language use in

response to the language spoken in the new context (A→B). The first possibility (A→A) is unlikely to arise, since one's language use constantly shifts in response to exposure (e.g. picking up new vocabulary). The fourth possibility (A→B) is also unlikely, as if one has acquired some new words in a language previously, it is unlikely that this usage will be entirely lost upon exposure to a new setting. This situation is applicable for those who have never learnt a particular language prior to exposure a new foreign language environment. The second (A→A-1) and the third (A→A+1) possibilities are more likely, as is a combination of these two. One may find that what s/he was taught previously cannot work effectively in a new environment, so s/he may discard arrangements of words that do not work (A→A-1) and adopt new ones (A→A+1). In this sense, what s/he has been taught previously changes and develops (A→A').

The second factor relates to the diversity of school education. It can be as minor as a situation in which each student is taught using the same materials but by different teachers. They may receive the same language input, but the way it is presented may result in some differences. Or, it could be more significant, such as when students are taught using different materials and by different teachers resulting in a greater probability that differences will arise. Some possibilities arise when we consider the diversity of school education, which may variously influence language learners. Taking EFL learners as an example, their language input is mainly from their teachers and materials provided by schools. Certainly, teachers in most EFL situations are the main sources of spoken input, and may teach English in their native languages, or half in their native language and half in English, or completely in English. The first is unlikely, since it is necessary to use English to teach English. The second and third options are relatively common in EFL classrooms. As regards materials, students may complete English courses in which they learn how to use English in general. In this case, English is the subject.

Other groups of EFL learners, may be moving on from learning English as a subject to using English as a study tool. These students may be learning some subjects that are non-English relevant, such as Chinese, Medicine, Engineering, using complementary materials written in English. In this case, English becomes the medium of learning. The opportunities to apply English when studying other subject expands the influence of English beyond what is taught in general English textbooks. There are also some subjects that are English relevant, for instance, English Literature or TESOL. In such cases, English is a medium and also a subject. Students of English-relevant subjects may have more awareness of how English works. In comparison to those learning non-language related subjects, EFL students would have even more opportunities to be influenced by their use of English.

Teachers / Materials	Completely in the native language	Half native language and half English	Completely in English
English (General)	(a) Unlikely to happen	(b) English as a subject	(c) English as a subject
Non English-relevant subjects	(d) English may still be the medium	(e) English as a medium	(f) English as a medium
English-relevant subjects	(g) Unlikely to happen	(h) English as a medium	(i) English as a medium

Table 3.1 EFL learning situations: English as a subject/medium

The table above presents all possible situations that might occur. Based on what has been discussed, situations (a) and (g) are unlikely. Also, it is unhelpful to include situation (d) if the aim is to see the influence on EFL learners' English. Furthermore, we can reasonably assume that every EFL learner has received general English courses before they move on to use English as a medium for their further studies. They have all experienced at least situation (b) or (c) previously. During a specified period, situation (c) (being taught in English/English is used as a medium in studying) might provide more language input than situation (b), but the probability is that EFL learners might differ greatly in terms of their exposure to course materials. When situations (e), (f), (h), and (i) occur, English is used as the medium of study. The students receive more varied English input in these situations than they do general English course materials. Therefore, it will be more likely for them to be influenced by input on their use of words or combinations of words in English.

We can classify the language input EFL students receive in situations (d) to (i) as shown in the following tables. The input could be all in their native language, partly in their native language and partly in English or all English (both spoken and written). For EFL learners studying non-English relevant subjects, if the language they receive at school is as described in situation (d1), it is possible that their English will remain the same as that taught in their general English courses. If they experience situations (d2) and (d3), they may change how they use English in response to written English input. Additionally, EFL learners in situation (d3) are likely to be influenced more than those in (d2) because they receive more written input. As for situations (e1) and (f1), it is hard to imagine in what context the EFL learners would have materials written in their native language but taught in English. However, if this arise, the EFL learners in situations (e1) and (f1) are likely to be influenced by spoken input and then their use of English will be influenced accordingly. In situations (e2), (e3), (f2) and

(f3), EFL learners can receive input from both spoken English and written English. Therefore, their use of written and spoken English is likely to be influenced as a consequence. Especially situation (f3) is most likely to affect learners' English use, since it provides the most abundant input under this condition.

Non English-relevant subjects

Written \ Spoken	All Native Language	Partly Native Language/Partly English	All English
All Native Language	(d1) unlikely to influence EFL learners' English use	(e1) likely to influence EFL learners' English use because of spoken input	(f1) likely to influence EFL learners' English use because of spoken input (more than e1)
Partly Native Language/Partly English	(d2) likely to influence EFL learners' English use because of written input	(e2) likely to influence EFL learners' English use because of spoken and written input	(f2) likely to influence EFL learners' English use because of spoken and written input (more in spoken English)
All English	(d3) likely to influence EFL learners' English use because of written input (more than d2)	(e3) likely to influence EFL learners' English use because of spoken and written input (more in written English)	(f3) likely to influence EFL learners' English use because of spoken and written input (potentially has the greatest influence)

Table 3.2 The EFL learning situations: English input in non-English-relevant subjects

As to the EFL learners who study in English relevant subjects (e.g. TESOL, English literature), the situations (g1) to (g3), (h1) and (i1) are not likely to happen (see Table 3.3), because it is impossible not to involve written English or spoken English for studying such subjects. Therefore, we can assume that these EFL learners are likely under the situations (h2), (h3), (i2) and (i3). In these situations (h2), (h3), and (i2), the EFL learners use both native language and (both written and spoken) English for studying, so their use of English is likely to be influenced by their English input. Since in the situation (h3), the EFL learners receive more written English input, their use of written English is likely to be influenced more. Same thing applies to the situation (i2). In the situation (i2) the EFL learners receive more spoken English input, therefore their use of spoken English is likely to be influenced. In the

situation (i3), the EFL learners receive the most abundant written and spoken English input. Therefore, we may assume that the influences on their English use are like to be greater in comparison to those students in the situations (d1) to (d3), (e1) and (f1).

English-relevant subjects

Written \ Spoken	All Native Language	Partly Native Language/Partly English	All English
All Native Language	(g1) unlikely to happen	(h1) unlikely to happen	(i1) unlikely to happen
Partly Native Language/Partly English	(g2) unlikely to happen	(h2) likely to influence EFL learners' English use because of spoken and written input	(i2) likely to influence EFL learners' English use because of spoken and written input (more in spoken English)
All English	(g3) unlikely to happen	(h3) likely to influence EFL learners' English use because of spoken and written input (more in written English)	(i3) likely to influence EFL learners' English use because of spoken and written input (potentially has the greatest influence)

Table 3.3 The EFL learning situations: English input in English-relevant subjects

This study explores whether both factors, the immersion environment and the diversity of school education, really have a significant influence on people's language use. Firstly, I would like to observe a group of people who had received similar language input in a similar environment. When they move on to experience different environments, or to receive a diversity of school education, do these factors really affect their use of words or combination of words? And if so how?

To observe the influences of these two factors on EFL learners' language use, I collected an EFL learner corpus. To ensure the results are valid, and to show the degree of relation to these two factors, I aimed to minimise the differences between the EFL learners' background variables (for example, socioeconomic status or educational level). My research participants were all EFL learners from Taiwan. They were born and grew up in Taiwan, and

had completed formal compulsory school education in Taiwan, and then undertaken senior high school and undergraduate studies in Taiwan. When they went on to attend graduate schools, their language environments and English input from courses changed. Some studied in Taiwanese universities, where English is not widely used for classroom teaching as a medium for students of non-English-relevant subjects, and some were students of English-relevant subjects at university on Taiwan using English as the medium of instruction. Others went to the United Kingdom to complete their postgraduate studies, where English was the medium for their studying and daily communication.

The reason for selecting Taiwanese postgraduate students as my participants was that Taiwan, according to Kachru's definition (1985, 2006), belongs to the Expanding-Circle. In the Expanding-Circle, English is a foreign language, and it is not commonly used in people's daily lives. The acquisition of English mostly arises from formal school education. Many people in Taiwan seek out additional English courses at their own expense. Institutes offering such courses are commonly known as '补习班(bǔxíbān)' or 'cram schools' in Taiwan. There are many types of cram schools available. Some follow the model of after-class schools, intended to reinforce input from attendees' regular English classes. This reinforcement is designed to improve learners' success in school examinations (the mid-term and final-term exams, or entrance exams for high school or university). Some are simply language schools which are not intended to be exam-oriented. These provide some courses to advance students' English proficiency, or offer English for special purposes (for example, business English). Regardless of the additional exposure available from this kind of courses, outside the classroom, English is not commonly used in learners' daily lives. In Taiwan, the majority of people use Mandarin Chinese and Taiwanese Southern Min. Their opportunities to use English are infrequent, and although access to English news, TV series and movies is relatively common, Chinese translations and subtitles are usually provided. It is relevant to note that additional English courses and English media might affect, to some extent, the Taiwanese EFL learners' English use. However, this kind of language input is not high frequency. It has some influence, but formal school education is considered to be the major element influencing Taiwanese EFL learners' perceptions of English use.

The fact that Taiwanese EFL learners lack sufficient English exposure outside the classroom makes them an ideal group to observe in terms of environmental factors and the influence of school education. To examine the impact of being in an immersion environment, the study includes a group of Taiwanese EFL learners undertaking postgraduate studies in the UK. All the participants were taking Master's degree programmes, after completing

undergraduate degrees in Taiwan. These Taiwanese EFL learners had lived in an English-speaking environment for more than one year. In terms of the second factor, the diversity of school education, the Taiwanese EFL learners in the UK were either in situations (f3) or (i3), and so it was assumed they were the group most likely to experience an influence on their language. The Taiwanese EFL learners were studying in Taiwanese universities, and to manage the environment factor I excluded potential participants with overseas travel experience. Thus, this group of participants had never visited an English-speaking country for more than three months. In fact, many of the participants had only travelled overseas for a short trip, and many had never been abroad. Therefore, their language use can be contrasted with that of the group studying in the UK to test the environmental factor.

The Taiwanese participants studying at a university in Taiwan universities also proved an ideal group for researching the second factor, the diversity of school education. There were two groups of participants: one studying in English relevant subjects (mostly TESOL), and the other studying non-English relevant subjects. The first group were likely to be in the situations (h2), (h3), (i2) or (i3). Based on my personal experience and conversations with my participants, situations (h3) and (i3) were judged the closest to authentic classroom situations. (The materials were written in English, and the teaching was either in English or in both English and Mandarin.) Additionally, based on my understanding of the context of university education in Taiwan, and conversations with my participants, the non-English-relevant group were found to be likely to be in situations (d1), (d2), (d3), (e2) or (e3). (For non-English relevant subjects, some subjects required the use of English materials (e.g. textbooks and other readings), but it was commonplace for teachers to use the native language at all times.)

<table>
<tr><th>Spoken
Written</th><th>All Native Language</th><th colspan="2">Partly Native Language/Partly English</th><th colspan="2">All English</th></tr>
<tr><td>All Native Language</td><td>(d1)/(g1)</td><td colspan="2">(e1)/(h1)</td><td colspan="2">(f1)/(i1)</td></tr>
<tr><td>Partly Native Language/Partly English</td><td>(d2)/(g2)</td><td colspan="2">(e2)/(h2)
[TW group: non-EN-relevant subjects]</td><td colspan="2">(f2)/(i2)</td></tr>
<tr><td>All English</td><td>(d3)/(g3)</td><td>(e3)/(h3)</td><td>[TW group: EN-relevant subjects]</td><td>(f3)/(i3)</td><td>[UK group]</td></tr>
</table>

Table 3.4 EFL learning situations: English input in different groups

Based on the information detailed above, I made two assumptions regarding immersion in the target language and the diversity of school education. The first was that the group of participants who went to the UK to study would have noticeable differences in their English use when compared to the group in Taiwan, because they had received more English input (and more diverse input). If input can really influence EFL learners' use of English, we should be able to observe significant differences. The second assumption was that, because of the diversity of school education (at the undergraduate and postgraduate stage), the English use of EFL learners studying in English relevant subjects would be noticeably different to that of those studying in non-English-relevant subjects. The input from classroom teaching (spoken English) and course materials (written English) differed from English-relevant subjects and non-English-relevant subjects. If input can influence how EFL learners use English, we should be able to discern significant differences in the English use of EFL learners based on receipt of different kinds of English input. The group studying in the UK might also be expected to reveal potential language influences from different classroom input. One additional observation about the influence of input also emerged in this study. Thus, I was interested to learn whether the different input from these factors (immersion in different language environments and different school education) had influenced the learners' use of English. The participants were all taught Mandarin as their first language, and it is also of interest whether the diversity of second/foreign language experiences affected their use of their native language. The details of the data collection procedures are given in section 3.2 below.

3.2 Building a Taiwanese EFL learner corpus

To achieve the research aims, in 2013, I conducted a series of interviews in Taiwan and the UK. The interviews were to be 30 minutes long and the interviewees were asked to speak in both English and Mandarin. The interviewees were informed that the interview would relate to school education and language learning. There were two sets of questions prepared for the interviewees, the first set to be answered in English and the other in Mandarin.

To ensure the English and Mandarin materials would be comparable, I formulated the two sets of questions to be as similar as possible (see the complete questions in Tables 3.5 and 3.6.). Questions E1 and M1 ask the interviewees to share their feelings about a language (English) and a subject (classic Chinese). E2 and M2 require the interviewees to express their thoughts about learning English and classic Chinese. E2 asks why the interviewees chose to

learn English, while M2 asks them if and why (or why not) they find learning classic Chinese helps them write better essays. E3 and M3 ask the interviewees to evaluate the quality of the English courses and writing courses they are taking, and E4 and M4 ask the interviewees to consider whether English proficiency and writing ability are likely to affect the development of their future careers. E5 and M5 ask the interviewees to discuss what factors make learning difficult/easy and what kinds of activities/books they like/dislike. E6 and M6 ask the interviewees to share their previous experiences using English outside school and their experiences in competitions relevant to language learning, such as speech or writing competitions. (In Taiwan, such competitions are very commonplace in school education.) E7 and M7 ask the interviewees to give their opinions about subtitles on films and simplified Chinese characters. I then asked them to evaluate whether subtitles are necessary and if so, whether switching from traditional Chinese characters to simplified ones will be a good idea in Taiwan (people continue to use traditional Chinese characters in Taiwan; these look different from simplified characters.)

	Questions, Part 1
E1.	Do you like English? Why?
E2.	What are/were the reasons why you learned English?
E3.	How do you feel about learning English in Taiwan? Did you like how English was taught in your junior high school/senior high school?
E4.	Do you think English is likely to be important in your future career? Please tell me your reasons why or why not.
E5	Do you think learning English is/was easy?
	If so, why? (Good learning strategies? Good resources? Good teachers? Or other factors...)
	If not, why? (Grammar? Speaking? Listening? Or other factors...)
E6.	Have you ever found English useful or interesting? If so, please tell me more about it.
E7.	Have you ever watched any English films without subtitles? How do you feel about a film without subtitles?

Table 3.5 English Interview Questions

These questions elicited illustrative and evaluative language use in response. In the process of interviewing, some follow-up questions were asked as applicable. For example, when an interviewee mentioned a good teacher who s/he had encountered before, I asked why s/he believed the teacher was good. When the interviewees spoke about the bǔxíbān/cram schools in Taiwan, I queried how they felt about the bǔxíbān/cram schools and the bǔxíbān/cram school teachers. Questions E1 to E7 and M1 to M7 were compulsory questions that every interviewee answered, but some additional questions were also asked during the interviews.

	Questions, Part 2
M1.	你喜欢学文言文吗? 为什么? (Did you like to learn classical Chinese? Why?)
M2.	有人说学古文对学生习写作文有帮助，你觉得呢? (Some people say learning classical Chinese is beneficial to students in writing. Do you think so?)
M3.	你对写作有什么想法? 你喜欢国中/高中时的作文课吗? (How do you think of writing an essay? Did you like the writing classes in high schools?)
M4.	你认为写作的能力对你的未来生涯重要吗？请说明你的理由. (Do you think writing is important in your future career? Please tell me your reasons why or why not.)
M5.	你喜欢阅读书籍吗？(Do you like reading?)
	如果喜欢，请告诉我你觉得哪类的书最有趣. (If YES, please tell me what kinds of books are most interesting to you.)
	如果不喜欢，请告诉我你觉得哪种活动比阅读更有趣. (If NO, please tell me what kinds of activities are much more interesting than reading.)
M6.	在你以前的求学过程中，你曾参加过任何和语言有关的竞赛活动吗? 如果有，请告诉我那是怎么样的竞赛活动. (Have you ever taken part in any language relevant competitions? If YES, please tell me what the competition was.)
M7.	你曾读过简体字吗? 请问你对简体字有什么看法? (Have you ever read anything in simplified Chinese characters? How do you feel about simplified Chinese characters?)

Table 3.6 Mandarin Chinese Interview Questions

The interviewees were informed that it was their comments on the issues of language learning and school education that mattered in this study. They were not informed that their use of language in the interviews would be a key area of investigation. The reason for slightly

misleading them in this was the concerns that the effects of their awareness on their language use might make it less natural. If the interviewees had been informed that their use of language, such as their lexical choices, collocations and grammatical structures were going to be examined in the research, their language use might have been unnatural because of their desire to be 'correct'. By shifting their focus onto other aspects of language, the chance that their natural language use could be achieved more easily was increased. Moreover, I encouraged the participants to ignore the need for 'correctness' when they spoke. Although the interviews took place in a formal setting, I did try to make the conversation as stress-free as possible. Although the spoken language used in the interviews was spontaneous and unscripted, it can be argued that the language used in the questions might have affected how these participants chose their words. It was also highly possible that my language use had some influence on their language although I tried not to speak too much during the interviews. Ultimately, collecting data via interviews was deemed the most practical and efficient mode of research. It made the data more comparable across the groups than random conversational materials would have done.

3.2.1 Three groups: LIVT, TESOL and NTESOL

There were two major groups of participants involved in the study. One was comprised of Taiwanese postgraduate students studying at the University of Liverpool, UK, and the other included Taiwanese postgraduate students studying in Taiwan (in 2013). Due to the circumstances of their recruitment, the participants in Taiwan were from different universities and colleges in northern Taiwan, mainly located in Taipei. There were three participants from a university in Taoyuan and one from Hsinchu. The second requirement was that all the participants were to have received formal school education in English in Taiwan for at least eight years. Their English competence was also controlled for. Their English levels fell between Intermediate to Upper-intermediate in General English Proficiency Test (GEPT) (equals to Rank 5.5 to 7.0 in IELTS). GEPT is a popular English proficiency test commissioned by Taiwan's Ministry of Education. To add a further background control, the participants recruited in Taiwan were required not to have been to any English-speaking countries for more than three months, and the participants recruited in the UK needed to have remained in the UK for at least one year. The length of stay was a key variable in this study. Ideally, I would have recruited people in Taiwan who had never been to any English-speaking countries, but this proved impossible. Therefore, I set three months as the maximum time that a participant in the second group could have spent abroad to be included, because someone who visits a country for three months can be regarded as a tourist and would probably have

held a tourist visa. A stay under three months would also be expected to influence language use less than a one-year stay.

ID	Gender	Discipline	Travel Experiences	English Level (GEPT)
TE01	F	Psychology (BA in TESOL)	Temporary stay (USA 3ms volunteer)	Upper-Intermediate (IBT 90)
TE02	F	Education (BA in TESOL)	Temporary stay (USA 5wks; UK 10dys)	Upper-Intermediate (IELTS 7)
TE03	F	TESOL	Short trips (Korea; Thailand)	Upper-Intermediate (ITP 570)
TE04	F	TESOL	Temporary stay non-EN (Turkey 2ms volunteer)	Upper-Intermediate
TE05	F	TESOL	Short trips (Thailand) (UK 8ms at the age of 7)	Upper-Intermediate
TE06	F	TESOL	None	Upper-Intermediate
TE07	F	TESOL	Short trips (Japan; UK)	Upper-Intermediate
TE08	F	TESOL	Short trips (China, Thailand)	Upper-Intermediate
TE09	F	TESOL	Short trips (Japan; China)	Upper-Intermediate
TE10	F	TESOL	Temporary stay (USA 2ms)	Upper-Intermediate
TE11	F	TESOL	None	Upper-Intermediate (TOEIC 750)
TE12	F	TESOL	None	Upper-Intermediate
TE13	F	TESOL	Short trips (Japan; Thailand; Indonesia)	Upper-Intermediate
TE14	F	TESOL	Short trips (HK; Australia)	Upper-Intermediate
TE15	M	TESOL	Temporary stay (Canada 3ms exchange student)	Upper-Intermediate (TOEIC 890)

Table 3.7 Background information about the TESOL participants

The number of participants totalled fifty: thirty-five people in Taiwan and fifteen people in the UK. The thirty-five people recruited in Taiwan were five males and thirty females. Their ages ranged from 24 to 32. It is important to mention here that people studying English Literature or TESOL may have greater linguistic awareness in English, so their performance may differ from those who had less awareness; consequently, the participants with English relevant backgrounds were treated as one group due to their specialty. There were thirteen participants studying TESOL at postgraduate schools and further two participants (TE01 and TE02) who had studied TESOL as part of their bachelor degree. These fifteen people were comprised one group (TESOL Group). (See Table 3.7.) There were three TESOL participants

who had never travelled abroad. Four TESOL participants had stayed in other countries for more than one month but for less than three months. Three of these had been to English-speaking countries, such as Canada, the UK and the USA, and one had been to Turkey, where English would presumably have been the lingua franca. Seven TESOL participants had experienced short trips to other countries, and the duration of their trips was less than one month.

One of these seven people, TE05, informed me that she had lived in the UK for eight months at the age of seven with her parents who were visiting scholars. She also told me that during that period she did not speak any English. She was sent to a primary school and spent her time with other children who could not speak English, and so she used her body language and gestures to communicate with people. I asked her whether she learned how to speak English by the end of her stay. She said that she did not learn much. Her memory of that period was a blur. She started to learn English formally when she was older, and did not have many opportunities to use English in her daily life after that time. Despite her early experience, the way she experienced English education at school and her lack of English use in daily life were comparable to the rest of the Taiwanese students. Therefore, I did not exclude TE05 from the research.

Another nineteen participants in Taiwan were placed in a different group, comprising those without a study background in either TESOL or English Literature. I labelled them the NTESOL group (non-TESOL). (See Table 3.8.) They were from various areas. Veterinary Medicine students were the largest sub-group among these. There were three TCSOL (Teachers of Chinese to Speakers of Other Language) students in the NTESOL group. Although they could be considered as having a linguistic-relevant background, they were not included in the TESOL group because the contents of their study differed from those of the TESOL group, and focused mostly on Mandarin. These students might also study English as it is the lingua franca for international students. However, their linguistic awareness of English was unlikely to be as comprehensible as that of the TESOL students.

One might be concerned that different disciplines would be an issue in terms of comparability. Indeed, from the interviews, it emerged that people with different research backgrounds expressed different opinions about the same topic. However, the key objective was to discover how individuals use words, instead of discussing ideas about the issues asked about in the interviews. Therefore, although differences might occur, the significance of these differences might not be serious enough to influence the results. This was why I placed the non-TESOL participants together as a group, instead of dividing the group. Initially, there

were twenty participants in the NTESOL group, although one of the participants told me in the interview that while the subject of her postgraduate study was Tourism, she had studied English literature as an undergraduate. It was difficult to place her in the TESOL group because the participants were all studying or had previously studied TESOL. It was not suitable to put her in the NTESOL group, because she might have received more English input than the other NTESOL participants. Due to these concerns, I excluded her from the study. Nineteen people remained in the NTESOL group. There were five NTESOL participants who had never been abroad, and four who had stayed in English-speaking countries for more than one month but for fewer than three months. There were ten NTESOL participants who had been to other countries for short trips.

There were fifteen Taiwanese participants in Liverpool who were placed in a group labelled the LIVT group. (See Table 3.9.) They had all stayed in the UK for one year or more for their postgraduate studies. Similar to the NTESOL group, the subjects of their studies were various. Notably, one LIVT participant, LIVT01, was studying applied linguistics, and LIVT05, had studied English literature in Taiwan. I did not separate these two from the other LIVT participants because of their linguistic-relevant background. This was because the purpose of having the LIVT group in the study was to collect a spoken language sample from people who had been in an English-based environment for a certain period. LIVT01 and LIVT05 met this requirement, so belonged in the group. However, to determine whether their linguistic-relevant background differentiated them from the other participants, these participants were viewed with extra care when investigating the data.

ID	Gender	Discipline	Travel Experiences	English Level (GEPT)
NTE01	F	Chinese	Short trips (Thailand; Singapore)	Intermediate (TOEIC 650)
NTE02	F	T Chinese SOL	Short trips (Thailand)	Upper-Intermediate
NTE03	F	T Chinese SOL	Short trips (South Asia)	Intermediate (TOEIC 440)
NTE04	F	Education	Short trips (Indonesia; Singapore)	Intermediate
NTE05	F	Education	Short trips (Vietnam; Korea)	Upper-Intermediate (TOEIC 770)
NTE06	F	Education	Temporary stay (USA 1m)	Upper-Intermediate
NTE07	F	Information Management	None	Intermediate
NTE08	F	Information Management	Temporary stay (UK 2ms language school)	Upper-Intermediate (IELTS 6)
NTE09	F	Law	Short trips (USA; Canada; Italy)	Intermediate (TOEIC 675)
NTE10	F	Veterinary Medicine	Short trips (Japan; UK; USA)	Upper-Intermediate
NTE11	F	Veterinary Medicine	None	Intermediate
NTE12	F	Veterinary Medicine	Short trips (Japan; South Asia)	Upper-Intermediate
NTE13	F	Veterinary Medicine	Temporary stay (USA 3ms Working holiday)	Upper-Intermediate
NTE14	F	Veterinary Medicine	None	Intermediate
NTE15	F	Veterinary Medicine	None	Intermediate
NTE16	M	Civic Design	Short trips (Thailand)	Upper-Intermediate
NTE17	M	T Chinese SOL	Temporary stay (USA 1m)	Upper-Intermediate
NTE18	M	Engineering	None	Intermediate
NTE19	M	Law	Short trips (Japan)	Intermediate (TOEIC 690)

Table 3.8 Background information about the NTESOL participants

ID	Gender	Discipline	Studying in the UK for…	Travel Experiences (outside TW/UK)
LIVT01	F	Applied Linguistics	1 year 10 weeks	Less than 3 months
LIVT02	F	Ancient History	1 year	Less than 3 months
LIVT03	F	Consumer Marketing	1 year 19 weeks	Short trip
LIVT04	F	Consumer Marketing	1 year 6 weeks	Short trip
LIVT05	F	Consumer Marketing (English literature in TW)	1 year 6 weeks	USA 1m/CAN 1m
LIVT06	F	Consumer Marketing	1 year 10 weeks	CAN 6mths
LIVT07	F	Management	1 year 6 weeks	Less than 3 months
LIVT08	F	International Business	1 year 6 weeks	Less than 3 months
LIVT09	F	International Business	1 year	Short trip
LIVT10	F	Advanced Management System & Technology	1 year 10 weeks	Over 3mths; less than 2yrs
LIVT11	M	Advanced Management System & Technology	1 year 6 weeks	Short trip
LIVT12	M	Advanced Management System & Technology	1 year 10 weeks	Australia 9mths
LIVT13	M	E-Business Strategy & Systems	1 year 6 weeks	Short trip
LIVT14	M	International Business	1 year 6 weeks	Short trip
LIVT15	M	International Business	1 year 6 weeks	Less than 3 months

Table 3.9 Background information about the LIVT participants

As for the LIVT participants' travel experiences, I used an online questionnaire to collect details of their previous travel experiences (outside the UK and Taiwan). They were asked to select from a list, with the aim of establishing which description fitted their situation best. They also had to report if they had been to other English-speaking countries. If they had visited non-English-speaking countries, they did not need to mention this. The option 'less than 3 months' indicates that they had been to the other countries for more than one month but fewer than three months. The option 'short trip' indicates that the length of the trip was less than one month. 10 out of 15 participants belonged to these two categories. There was one LIVT participant, LIVT10, who selected the option 'over 3 months, but less than 2 years' but did not mention the country, but I learned in the interview that it was China. Another two participants said they had stayed in other English-speaking countries for a long time. LIVT05

had stayed in the USA and Canada for one month each. LIVT06 had been in Canada for six months, and LIVT12 had stayed in Australia for nine months. I had some concerns about whether LIVT10's extended experience travelling in China would influence her Mandarin use, and whether in the cases of LIVT05, LIVT06 and LIVT 12, their additional experience in other English-speaking countries would influence their language use. Nevertheless, since the main purpose of creating the LIVT group was to establish whether a long period of stay in an English-based environment would influence participants' language use, I kept these participants in the LIVT group. However, similar to the way participants with extra linguistic study experience were treated, the participants with extra travel experience were identified and paid additional attention to in the analysis.

Since the Taiwanese participants were studying in the UK, they were unquestionably exposed to (written/spoken) English input in class. The input could be during lectures, tutorials and group meetings for school projects. In comparison with the Taiwanese participants studying in Taiwan, these LIVT participants were more likely to receive spoken/written English (input), and to produce spoken/written English (output). The LIVT participants lived in Liverpool, which is an English-speaking environment, so the language input outside their classroom should also be considered. When asked how much time the LIVT participants usually spend using English to communicate each day, 7 out of 15 participants (47%) admitted using English for less than 1 hour after school on average. 4 out of 15 LIVT participants (27%) stated that the average time spent communicating in English after school was 1 to 2 hours each day, and 3 (20%) said the average hours they spent communicating in English outside the classroom was 2 to 3 hours per day. Only one LIVT participant (6%) claimed she used English to communicate more than 3 hours a day after school (see Table 3.10).

	(Outside Classroom) How much time do you usually use English for communication per day?			
N	Less than 1 hour	1 to 2 hours	2 to 3 hours	More than 3 hours
1	LIVT04	LIVT01	LIVT05	LIVT06
2	LIVT07	LIVT02	LIVT11	
3	LIVT08	LIVT03	LIVT12	
4	LIVT09	LIVT10		
5	LIVT13			
6	LIVT14			
7	LIVT15			

Table 3.10 Hours of English use after school (LIVT participants)

When asked with whom the LIVT participants used English after class, the answers included roommates/flatmates, clerks, co-workers, customers and friends who did not speak Mandarin Chinese. We found almost half of our LIVT participants seldom spoke in English after class, and this was because they were either living alone or with other Mandarin speakers. The participants who used English more than 1 hour per day for communication often lived with other international students. The participant who used English more than three hours each day outside class had a part-time job.

When asked what sorts of extracurricular English input the LIVT participants usually had, most reported watching TV/movies/videos (e.g. YouTube) or listening to the radio regularly. Four LIVT participants admitted that they were not exposed to any particular English input after class (see Table 3.11).

	Extracurricular English Input
LIVT01	Radio/ TV (1-2 hours a day)
LIVT02	TV (1 hour a day)
LIVT03	TV (4-5 hours a day)
LIVT04	Radio (1 hour a day)
LIVT05	TV (4-5 hours a day)
LIVT06	Movie/Video (sometimes)
LIVT07	TV (1-2 hours a day)
LIVT08	No
LIVT09	Movie/Video (1 hour a day)
LIVT10	No
LIVT11	No
LIVT12	Movie/Video (1.5 hour a day)
LIVT13	Radio/Video (sometimes)
LIVT14	No
LIVT15	Radio (1-2 hours a day)

Table 3.11 Hours of English use after school (LIVT participants)

It emerged from the data that even though Taiwanese postgraduate students were studying in the UK, the amount of English input they received, and the degree of their immersion in English varied. Therefore, it should be underscored that although these participants were considered as one group, their experiences were different. However, despite the differences in their use of extracurricular English, all LIVT participants were taking courses which required them to use English. In addition, all the LIVT participants were in the

final stage of their Master's programmes, and were either writing their dissertations or had finished doing so. This distinguishes them from the Taiwanese postgraduate students studying in Taiwan, who were either using a limited amount of English for their studies or none at all. Moreover, the TESOL and NTESOL participants rarely had opportunities to use English after school in Taiwan. Thus, it is still possible to investigate the potential influences from language immersion in the LIVT group by comparing their language with that of the two Taiwanese groups.

3.3 Method of compilation

In corpus linguistic studies, there are mainly two types of corpus comparison: (A) '*comparison of a sample corpus with a large(r) standard corpus*'; (B) '*comparison of two (roughly-) equal sized corpora*.' Researchers use the first type of comparison (A) to discover features of language in a research corpus (e.g frequency) that differ significantly from general language (large standard corpus, e.g. BNC or COCA). The second type of comparison (B) aims to identify features of language which distinguish one corpus from another corpus. (Rayson, Berridge and Francis, 2004). In this study, we intend to compare language use across different EFL learners' corpora (LIVT, TESOL and NTESOL groups) to ascertain whether different variables (language immersion/different school education) generate noticeable differences in language use. Therefore, the main comparison in the current study is the second type of comparison; i.e. a comparison of three similar sized corpora. If a particular feature of language use is found in these three EFL learner corpora that seems to be very different from general language, then an initial comparison (A) will be made for examination purposes. Therefore, in this study, we need to compile three roughly equal sized EFL learner corpora and to use a general corpus if necessary. As illustrated in 3.2, we collated language use data for three different EFL learner groups in interviews. Section 3.3.1 describes how these materials were transcribed to create the corpus data.

3.3.1 Data Transcription

I transcribed all the interviews for consistency, using the CLAN program, developed by Brian MacWhinney (2000) and his team. The CLAN program is a corpus tool developed to study spoken language (children's spoken language mostly). Its developers suggest that users follow norms when tagging and labelling transcriptions, such as labelling repetition and overlapped conversation. I did not, however, follow these transcription rules exactly, because I wanted to observe the collocation of words in this study, and so recording discourse behaviours such as repetition or turn taking was not a concern. Since this study principally examines collocations, I did not direct attention to the issue of phonology, either. I decided to ignore the 'non-standard' pronunciation of words that participants produced, and transcribed words in common spellings/characters. These 'non-standard' pronunciations would be of great potential research value in future studies, since potential phonological differences also reveal how participants are influenced by language input. However, due to the limitations of space in the study, this area was not explored. Another issue that arose when transcribing was that some words were unfinished. The speakers did not complete words; I transcribed these with round brackets to indicate the missing parts of the words, e.g. *becau(se)*.

The same principles were applied to both English and Mandarin Chinese transcriptions. However, I required another corpus tool for the Mandarin Chinese transcriptions. The chosen tool was an online system; the CKIP Chinese Word Segmentation System (中文斷詞系統) (Ma and Chen, 2003), developed by Academia Sinica (or Central Research Academy) in Taiwan. It tokenises plain Mandarin texts into analysable texts. Unlike English, where Mandarin Chinese is presented in a written form, there are no spaces between words. Therefore, if researchers use corpus tools to analyse Mandarin, they then need to separate the Mandarin text into segments.

My intention was to study the language use of the respondents, so the data file does not contain my side of the dialogue. The extent of the EFL learners' English and Mandarin corpora are presented below.

English	LIVT (15)	TESOL (15)	NTESOL (19)
Tokens	28870	29685	23305
Types	1607	1932	1396
Standardised TTR	24.35	26.04	22.61

Table 3.12 The sizes of the LIVT, TESOL and NTESOL English corpora

Mandarin	LIVT (15)	TESOL (15)	NTESOL (19)
Tokens	38089	34772	41420
Types	2551	2855	3055
Standardised TTR	27.26	29.01	29.43

Table 3.13 The sizes of the LIVT, TESOL and NTESOL Mandarin Chinese corpora

3.3.2 Reference English Spoken Corpora

The study hypothesis was that the use of English among the LIVT group would show noticeable differences from usage by the other two Taiwanese groups (TESOL/NTESOL). The objective (if this is proven) would then be to ascertain whether the LIVT group's English use was influenced by input received in the UK. To determine this, a general British corpus was also needed, and so the BASE[1] (British Academic Spoken English) corpus, was selected as a reference corpus for the study (Thompson and Nesi, 2001). The BASE Corpus comprises lectures and seminars recorded at the University of Warwick and the University of Reading. It contains 1,644,942 tokens in total, and has four broad disciplinary groups. I only selected texts from the Arts and Humanities group, because the topic in the interviews was in the Arts and Humanities domain.

[1] The recordings and transcriptions used in this study come from the British Academic Spoken English (BASE) corpus. The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson. Corpus development was assisted by funding from BALEP, EURALEX, the British Academy and the Arts and Humanities Research Council.

A further reason existed for selecting BASE as the reference corpus, instead of another general English corpus (e.g. BNC); that is, the nature of BASE makes it similar to the input which the Taiwanese students had received. In Taiwan, English is a foreign language and there is little opportunity for ordinary Taiwanese people to hear and use English in their daily lives. Thus, the majority of their English language input is received in an academic context (at school). This also applies to the LIVT participants, as we discussed at the end of section 3.2, as the main English input they receive is in an academic context. If I had selected another general English spoken corpus, comprising casual conversations, this would be less comparable than the Taiwanese corpora.

The size of the reference corpus (BASEah) is shown as below:

	Token	Type
BASEah	431515	18366

Table 3.14 The size of the BASEah corpus

	Token	Type
COCAsp	18,452,848	81,396

Table 3.15 The size of the COCAsp corpus

Another English corpus was used as a general norm for reference purposes as in Taiwan, the major English input is from the USA, as US TV news and US TV programs are readily accessible there. In addition, many language learning materials are provided by American publishers. Therefore, using the general corpus, the COCA (Corpus of Contemporary American English) was used as a reference corpus for this study (Davies, 2008). The COCA contains spoken texts from TV news and interviews from 1990 to 2012, and materials from the last 5 years (2008-2012) were used as the general reference corpus for this thesis. This corpus was labelled COCAsp. Therefore, the main reference corpus was BASEah, but COCAsp was also used for the analysis where necessary.

3.3.3 Reference Mandarin Spoken Corpus

In terms of the use of Mandarin, I hypothesised that the EFL learners' Mandarin use would not differ greatly from that used by the other groups, since they learned their Mandarin Chinese as first language speakers in a natural Chinese-speaking environment. However, if any noticeable features of language use emerge as needing to be examined, a general Mandarin corpus will be consulted.

Finding an equivalent Mandarin comparator corpus to the BASE and COCA corpora was not possible, however, I found a suitable online corpus platform for researchers wishing to investigate Mandarin words. This is the Academia Sinica Balanced Corpus of Modern Chinese (Sinica Corpus), which includes spoken Mandarin materials and can be used for research purposes (Chen and Huang, 2015). It provides both written and spoken Mandarin sub-corpora in various genres. I utilised its conversation and interview materials in the analysis as necessary. The size of the Sinica corpus is as detailed below:

	Token	Type
Sinica Corpus (spoken)	575,500	379,400
Conversation/interviews	129,000	13,200

Table 3.16 The size of the Sinica (spoken) corpus

3.4 Method of Analysis

3.4.1 Words for Analysis

When we discussed language acquisition in Chapter 2, the frequency of language instances was considered a key element. When one pattern of stimuli and meanings/information occurs frequently, it is likely to be stored in a person's long-term memory. Language is 'built up' of multi-words, which can be joined ritualistically according to frequency and idiomaticity (Nattinger and DeCarrico, 1992; Lewis, 1993). Therefore, one's language use, to some extent, reflects the language instances one has encountered previously. If certain words/collocations of words/patterns occur frequently, they are likely to be stored in the acquirers' memory for subsequent use by acquirers. Consequently, when researchers want to investigate language use, frequency of language instances is usually the focus. To investigate the frequency of language instances, a quantitative analysis is required.

Quantitative analysis of a corpus requires findings "to be generalised to a larger population" and allows different corpora to be compared directly (McEnery and Wilson, 2001: 76). One of the approaches to studying quantitative data is "to classify items accordingly to a particular scheme and to perform an arithmetical count of the number of items within the text which belong to each classification within the scheme" (McEnery and Wilson, 2001: 82). In this study, the lexical items investigated were chosen according to frequency counts.

To ascertain frequency counts, the *WordList* function in Wordsmith Tools 6.0 (Scott, 2012) was used to prepare a list of words arranged according to their order of frequency. I selected the top 30 most frequent words from the list in the Taiwanese corpus of candidates (see Appendix 1 for top 30 frequent wordlists). The criteria for word selection were: first, that the words should be present among all three Taiwanese groups and in the reference corpora. Second, the words should appear in all three Taiwanese groups' top 30-word list. Third, the words of less interest considering the research goals would be neglected. The primary focus of this research was on the collocation of words, and I intended to establish how words collocate when composing a sentence. Therefore, those words that were too content-relevant or less meaningful were not suited to fulfil this aim. For instance, filler sounds, such as *eh*, *啊 (ā)* or *嗯 (èn)* were removed, because they had little meaning in the sentences. A similar situation applied to the nouns *English* and *school*. Both words appeared with a high frequency because the topic in the interviews concerned both learning English and school education. Their frequency counts were only high because they were the topics being discussed, so these words were not chosen for analysis. Fourth, if words appeared in both the English and Mandarin Chinese top-30-word lists then they should be selected with a high priority. In this study, although the primary focus is on potential differences in EFL learners' English use, whether Mandarin was used by EFL learners was also of interest. Therefore, if high frequency English words had nearly-equivalent Mandarin words with high frequency counts, then these English and Mandarin words were selected for analysis. According to these criteria, the pronouns *I, you, it* and the Mandarin pronouns *我 wǒ (I), 你 nǐ (you)* and *它 tā (it)* were chosen. The reason for analysing pronouns was that (1) there were adequate instances of them in our EFL learner corpora. One of the limitations of this study is the size of the sample corpora, as if instances are too few, it is difficult to generalise on the basis of them. (2) Pronouns usually have conjunctions as their L1 collocates, and are usually followed by adverbs or verbs. Therefore, it would be a useful starting point to investigate how these lexical items are used in data. We will investigate *I, you,* and *it* in Chapter 4 and *我 wǒ (I), 你 nǐ (you)* and *它 tā (it)* in Chapter 5.

Five candidate verbs for discussion were present in the English list. These were *think*, *have*, *can*, *is* and *like*. On the Mandarin list, there were eight candidate verbs: 觉得(juéde) (meaning 'to think or to feel'), 有(yǒu) (often meaning 'to have, to exist'), 会(huì) (often meaning be capable of doing things), 是(shì) (meaning 'to be'), 说(shuō) (meaning to say/speak), 写(xiě) (meaning to write), 看(kàn) (meaning to see/look) and 要(yào) (meaning to want). Therefore, I selected the English words *have* and *is* and the Mandarin words 有(yǒu) and 是(shì) for analysis because they were nearly-equivalent. The words 有(yǒu) and 是(shì) are commonly translated into English as *have* and *is* (*BE-verbs*). We will investigate *is* and 是(shì) in Chapter 6 and *have* and 有(yǒu) in Chapter 7.

3.4.2 Procedure of Analysis

In corpus research, a test of statistical significance is typically used to determine whether a difference in language use between two texts/corpora has arisen by chance. There are several significance tests available, e.g. the chi-squared test and the t-test; of these the chi-squared test is one of the most frequently applied (Oakes, 1998). The chi-squared test compares differences between the actual frequencies found in a corpus (called 'the observed frequencies) and the frequencies with which researchers expect words to occur if no factors were affecting the frequency of use (called 'the expected frequencies'). If the observed frequencies are close to expected frequencies, it probable that they are occurring by chance. If the variance between the observed frequencies and expected frequencies exceeds that predicted, it is possible that the observed frequencies are being influenced by external factors (not arising by chance) (McEnery and Wilson, 2001: 84-85). Although the chi-squared test is a powerful tool for testing significance, the chi-squared statistic becomes less reliable when the expected frequencies are small (Butler, 1985).

Therefore, Dunning (1993) proposes the log-likelihood ratio as an alternative test for smaller texts/corpora. Cressie and Read (1984) point out that the two statistical tests are "in fact two statistics in a continuum defined by the power-divergence family of statistics" (Rayson, Berridge and Francis, 2004: 929). The chi-squared test relies on the assumption that there is a normal distribution, and the log-likelihood ratio includes a parametric analysis that relates to binomial/multinomial distributions. Dunning (1993) therefore suggests that where texts/corpora are small, the log-likelihood ratio serves as a better tool to test significance. In their research, Rayson, Berridge and Francis (2004) validated this statement by using both the chi-squared test and the log-likelihood ratio to test various combinations of corpora.

Since the EFL learner corpus in this study is rather small, the log-likelihood ratio is the best suited test of significance. Other tools exist to help researchers evaluate the significance of collections; e.g. the mutual information score (MI score), and the Z-score. These two statistical tests assist researchers wishing to identify whether words found together (collocations) co-occur by chance. For instance, if pairs of words have highly positive MI scores, it is to be expected that they will have a special relationship to each other, rather than being paired randomly (McEnery and Wilson, 2001: 86). In this study, our principal aim is to compare language use across three EFL learner corpora to investigate whether our groups of EFL learners are using special words/word sequences with any degree of significance. The relationships between words are of less interest in the current study. Therefore, we will mainly use the log-likelihood ratio when conducting the data analysis.

In this study, WordSmith Tools 6.0 (Scott, 2012) was utilised for the purpose of analysis. The C*oncord* function in WordSmith Tools allows people to view a word or a phrase in context, and thus my analysis relies heavily on this function. Firstly, I searched for the target word and developed a list of concordance guidelines. Then, I checked the R1 and L1 collocates of the words and arranged them according to frequency. I extracted the top 10 most frequent R1 and L1 collocates of each word for comparison. This method was applied when analysing both English and Mandarin words to the Taiwanese corpora and the reference corpora, BASEah and COCAsp. I was unable to apply the same method with the Mandarin reference corpus, SINICA, because the data set could only be accessed via its web interface. However, it is possible to collect a list of the collocates of the target word in order of frequency via the web tool provided by SINICA, and so the use of different corpus tools did not influence the data collection. The concordance lines and results of the collocate lists were then used as materials to view the target word's collocation, colligation (grammatical features) and semantic associations. After the lists and figures were obtained, I used the Log-likelihood Ratio Calculator, created by Jiajin Xu (2009), to search for statistically significant differences in terms of frequencies across the groups. If collocations/grammatical features was found more /less frequently in one group, a further investigation on this usage was then conducted. I observed the raw and standardised frequencies (time/per 1000 words) for the language used, alongside examples from all the Taiwanese EFL learner groups. The instances could then be arranged based on their features, and the findings compared with the uses that emerged in the reference corpora. This was crucial to establish whether the differences supported our hypotheses (potentially influenced by two factors, language immersion and different school education).

Chapter 4: Use of Pronouns *I*, *You* and *It* by the three Taiwanese groups

4.1 Introduction

In chapter 2, I identified four hypotheses that this thesis would investigate. I repeat them here for convenience:

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

In this and the next chapter, we will investigate the use of the English pronouns *I*, *you* and *it* and the use of Mandarin pronouns 我 [I], 你 [you] and 它 [it]. The first person pronoun *I* is the most frequent word in the LIVT group and the TESOL group and the 2nd most frequent word in the NTESOL group. The first person pronoun 我 [I][2] is the 2nd most frequent word in all three groups. It is not surprising to find that the first person pronouns *I* and 我 [I] are frequently used by the three

[2] In Mandarin, the word 我 [I] is used in possessive, 我 [I] with grammatical particle 的 [possessive particle], and plural form, 我 [I] with grammatical particle 们 [plural marker for pronouns]. My Mandarin corpus was segmented with the Mandarin Chinese Tokenization service provided by SINICA. It treats the plural form 我们 [we] as one entry and the possessive form 我的 [my] as two entries 我 [I] and 的 [possessive particle]. Therefore, with the search word 我 [I], the word 的 [possessive particle] will be seen as its R1 collocate.

groups. After all, the theme of our interviews was to discuss their personal experience and beliefs. However, when we compare the use of first person pronouns in both English and Mandarin, we find that the use of Mandarin first person pronoun 我 [I] is used proportionally less often in all three groups than is the English first person pronoun *I*. This may suggest that the participants use such words in these two languages differently, even with a word which is as apparently uncomplicated as the first person pronoun.

		Rank	Freq.	Per 1000 Words
I	LIVT	1	1826	63.25
	TE	1	1915	64.51
	NTE	2	1657	71.10
		Rank	Freq.	Per 1000 Words
我 (wǒ)	LIVT	2	1711	**44.36**
	TE	2	1270	**35.96**
	NTE	2	1585	**37.63**

Table 4.1 Frequencies of *I* and 我 (wǒ) in LIVT, TESOL and NTESOL

The English second person pronoun *you* is the 5th most frequent word in the LIVT group, the 18th most frequent in the TESOL group and the 19th most frequent word in the NTESOL group. The Mandarin second person pronoun 你 [you] is the 4th most frequent word in the LIVT group, the 9th most frequent in the TESOL group and the 11th most frequent in the NTESOL group. Again, therefore, there might be something worth studying with regard to the frequency of second person pronouns *you* and 你 [you]. In particular, the LIVT group seems to have a substantial inclination to utilise the second person pronouns in their language use more than the other two groups.

		Rank	Freq.	Per 1000 Words
you	LIVT	5	714	**24.73**
	TE	18	299	10.07
	NTE	19	236	10.13
		Rank	Freq.	Per 1000 Words
你(nǐ)	LIVT	4	1014	**26.29**
	TE	9	632	17.90
	NTE	11	637	15.12

Table 4.2 Frequencies of *you* and 你 (nǐ) in LIVT, TESOL and NTESOL

The impersonal pronouns *it* and 它 [it][3] are also frequently used pronouns. As with *you* and 你 [you], there are some tendencies worth investigating. The NTESOL group shows a tendency to use the Mandarin impersonal pronoun 它 [it] more often than the other two groups.

IT (including *IT'S*)		Freq.	Per 1000 Words
	LIVT	699	24.21
	TE	630	21.22
	NTE	577	24.76
它(tā)		Freq.	Per 1000 Words
	LIVT	331	8.58
	TE	296	8.38
	NTE	515	**12.23**

Table 4.3 Frequencies of *it/it's* and 它 (tā) in LIVT, TESOL and NTESOL

In the light of the differences noted in the frequencies of these words, a further investigation seems warranted. In the next sections, we will discuss the use of English first person pronouns *I* in 4.2, and then proceed to investigate the English second person pronoun *you* in 4.3. The English impersonal pronouns *it* will be investigated in 4.4. The uses of Mandarin pronouns 我 [I], 你 [you] and 它 [it] will be considered in Chapter 5.

Before our investigation on the use of collocates, there is one thing needed to be stated. In our investigation of the noticeable differences between groups, we will try to identify whether this kind of difference is in the domain of discourse difference or is in the domain of linguistic difference (lexico-grammatical difference). The latter is a difference showing in the use of a collocation (co-occuring words)/colligation (grammatical features) and semantic association of a word or word sequences. When one particular use of collocation/colligation and semantic association of a word or word sequences is used only by one group but not (or significantly less) by the other groups, this is a

[3] Unlike English which distinguishes the third personal pronouns *he* and *she* clearly from the impersonal pronoun *it*, in Mandarin, the pronunciation of the third personal pronouns 他 [he] and 她 [she] is the same as the impersonal pronoun 它 [it]. I separated the impersonal pronoun 它 [it] from the third personal pronouns 他 [he] /她 [she] based on the content of speech while transcribing. The use of third person pronouns and impersonal pronouns are already separated in the corpus. Therefore, there is no overlapping of the third person pronouns 他 [he] /她 [she] and the impersonal pronoun 它 [it] in our analysis and discussion.

linguistic/lexico-grammatical difference. In such cases, we will try to see whether this difference is likely to have been influenced by the different English input that these participants have received.

The former kind of differences (discourse difference) is where one group of participants favours the use of particular discourse strategies more than the others. Filler sounds and repetitions of words are used by every participant. These are discourse strategies for speakers to obtain more time for generating their ideas, or to suggest that they have not yet finished their utterances. In such circumstances there is little difference between groups as regards the linguistic/lexico-grammatical aspect of this use. The main question regarding this kind of differences is why one group of participants used one particular discourse strategy significantly more or less often than the other groups when they were in similar contexts. All of the participants were interviewed individually under similar circumstances, in a room where only two people (the interviewer and interviewee) presented. There were no other people to hear the talk, so the interviewees could talk freely without affecting or being affected by the language of others (beside the interviewer). The interviewer was the same person all along. They were asked almost identical questions. Therefore, it is intriguing to see that one group of these individuals who shared one type of similar backgrounds behaved significantly differently from the other groups of individuals who had another type of similar backgrounds. However, although we can identify the differences between these participants' discourse preferences, the reason behind this kind of differences is not easy to be seen.

The differences in people's discourse behaviour may be affected by various factors. There may be social factors in which, for instance, the people typically around this individual would have an effect on him/her. If an individual is embedded in an environment where people tend to speak in a soft tone and do not encourage people to make a firm judgement or a strong claim, this individual is very likely to follow similar discourse strategies in his/her language use. It may also be the nature of the social interaction that affects the behaviour of this individual. For instance, the LIVT participants were more likely to see me (the interviewer) as a fellow student. The TESOL participants were likely to see me as either a more advanced student or a fellow student (since I was a TESOL student as well.) The NTESOL participants, however, might see me as a serious examiner/researcher, someone at distance from them socially. These different social roles may have affected the interaction and therefore the participants' discourse behaviour as well. There is yet another possible factor which might have affected the participants' discourse behaviour – the educational factor. The different discourse usages/preferences may be affected by the different educational backgrounds of individuals. For instance, one who has been trained for the study of science may put his/her focus on the existing facts and tends to illustrate things in an objective way. The different training that people have in their fields may affect their behaviour in discourse. In this study, we do not aim to find out which factor is functioning behind the noticeable discourse differences between these groups. If there is a likely explanation behind the difference of discourse behaviours, we will mention it in our discussion. Still,

our main concern for this study is to identify the different kinds of linguistic differences between the groups and try to see whether the input factor is a plausible cause of the differences.

4.2 The use of pronoun *I*

In our previous brief observation on the frequency of the pronouns *I* and 我 [I], we noted some differences between the use of English *I* and Mandarin 我 [I]. When we compare the use of *I* and 我 [I] via the log-likelihood ratios, we find that the NTESOL group employed the English pronoun *I* significantly more frequently than the LIVT and TESOL groups. With regard to its use of the Mandarin pronoun 我 [I], the LIVT group used 我 [I] significantly more frequently than the TESOL and NTESOL groups. (See Table 4.4.)

	LLR	Sig.		
LIVT-TE	0.365	0.546		-
LIVT-NTE	11.866	**0.001**	***	-
TE-NTE	8.381	**0.004**	**	-
	LLR	Sig.		
LIVT-TE	32.381	**0.000**	***	+
LIVT-NTE	22.304	**0.000**	***	+
TE-NTE	1.449	0.229		-

Table 4.4 Log-Likelihood Ratios: *I* and 我 (wǒ) in LIVT, TESOL and NTESOL

The findings suggest that although these participants were asked the same set of questions in English and Mandarin about their personal thoughts and experiences, the way these groups used the first person pronouns are quite different. In English, it shows greater use of English *I* in the NTESOL group than in the LIVT and TESOL groups. In Mandarin, the LIVT group made more use of Mandarin first person pronoun 我 [I] than the TESOL and NTESOL groups. The use of the first person pronoun in English is almost twice as frequent as in Mandarin. The findings that one group used significantly more personal pronouns than the other groups suggest that there are potentially some differences between these groups. The significantly high use of the first person pronoun *I* in the NTESOL group in comparison with the use in the other two groups may suggest that the NTESOL participants were more likely to concentrate on themselves in their English answers than the other two group participants. There are no significant differences in the frequencies of the use of *I* between the LIVT and TESOL groups. The finding that the LIVT and TESOL participants both used the pronoun *I* less in comparison with the NTESOL participants may suggest that, in their English answers, they tended to describe things from the viewpoint of a third party more often than the NTESOL participants did. It is comparatively more difficult for a EFL learner to give their opinions about an

unfamiliar or impersonal subject, since it may involve many unfamiliar vocabulary, complex noun phrases or sentence structures. It is comparatively easier for EFL learners to talk about themselves, since it is a familiar subject, and the vocabulary and the complexity of the sentence construction will be easier to handle. The NTESOL participants were the ones who had fewer opportunities to hear and use English in their daily lives and studies. It is possible that the NTESOL participants were less confident in talking about subjects other than themselves in English. The LIVT and TESOL participants, in contrast, were the ones who had more opportunities to hear and use English in their daily lives and studies. They could be more confident in talking about all sorts of subjects in English without fearing too much a lack of words or the complexity of constructing a sentence. In the use of Mandarin Chinese, the more frequent use of the first person pronoun 我 [I] in the LIVT group (or the lack of the use of 我 [I]) in the TESOL and NTESOL group) may suggest that the LIVT group used the pronoun 我 [I] differently from the way the TESOL and NTESOL groups did. (There are no significant differences between the TESOL and NTESOL groups in the frequencies of the use of 我.) It is possible that the LIVT participants talked about themselves more often than those in the TESOL and NTESOL groups in the interview. It is also possible that the LIVT participants did not omit the pronoun 我 [I] in their Mandarin as often as the TESOL and NTESOL participants did. In Mandarin, it is commonplace to omit the pronoun 我 [I] (subject or object) since both the speaker and hearer know who is doing the talk. As the example shows:

嗯	不	喜	欢	。	=	嗯	[我] 不 喜 欢 [它] 。
Èn	bú	xǐhuɑn				Èn	wǒ bú xǐhuɑn tā
Eh, [I] don’t like [it].						Eh, I don’t like it.	

Although it might have been interesting to investigate whether these assumptions are supported by an analysis of the Themes in their language use or how many omissions of 我 [I] there are in the groups, such a study would have fallen outside the main concern of my study. What is indisputable is that there is evidence to suggest that these groups used language differently. In the following section, we will therefore investigate the collocations of the use of *I*. The use of Mandarin first person pronoun 我 [I] will be looked at in the next chapter.

4.2.1 *L1 Collocates of I*

When we check the L1 collocates of the first person pronoun *I* in the three groups, we find that most of the top 10 most frequent L1 collocates are the same in the three groups. The filler sound *eh* and the repetition of the pronoun *I* are the most frequent collocates. Likewise, the connectives *so, and, when, because, but* and *if* are also frequently used in conjunction with *I* by the three groups. The words *yeah, yes,* and *think* only appear in one or two groups' top 10 most frequent L1 collocate lists. (See Table 4.5.)

	LIVT	1826	63.25	**TE**	1915	64.51	**NTE**	1657	71.10
N	**L1**	**Freq.**	**Per 1000 words**	**L1**	**Freq.**	**Per 1000 words**	**L1**	**Freq.**	**Per 1000 words**
1	EH	183	6.34	EH	198	6.67	EH	248	10.64
2	SO	142	4.92	SO	178	6.00	I	190	8.15
3	I	138	4.78	I	148	4.99	SO	136	5.84
4	WHEN	101	3.50	WHEN	130	4.38	AND	92	3.95
5	BECAUSE	72	2.49	AND	120	4.04	BECAUSE	91	3.90
6	AND	71	2.46	BECAUSE	86	2.90	WHEN	69	2.96
7	BUT	67	2.32	BUT	81	2.73	BUT	59	2.53
8	**YEAH**	**47**	**1.63**	IF	38	1.28	IF	44	1.89
9	IF	45	1.56	**YES**	**35**	**1.18**	**YES**	**39**	**1.67**
10	THINK	42	1.45	**YEAH**	**32**	**1.08**	THINK	29	1.24
--	**YES**	**23**	**0.80**				**YEAH**	**23**	**0.99**

Table 4.5 L1 collocates of *I* in LIVT, TESOL and NTESOL

Some use of collocations showed significance. (See Table 4.6.) We can see that the NTESOL participants used the filler sound *eh* before the pronoun *I* and the repetition of the pronoun *I* (*I I*) significantly more frequently than the LIVT and TESOL participants. It shows no significant differences in the frequencies of these two usages in the LIVT and TESOL groups. Frequent use of the filler sound *eh* with *I* and the repetition of *I* indicates hesitation in speaking. It is a common behaviour observed in all the groups. The LIVT participants used the filler sound *eh* with the pronoun *you* and the repetition of *you* significantly more frequently than those in the TESOL and NTESOL groups in 4.3. The significantly more frequent use of *eh I* and *I I* in the NTESOL group is possibly caused by the fact that the NTESOL participants used the first person pronoun *I* significantly more often than the ones in the other two groups as we had discussed above. The highly frequent use of *eh you* and *you you* in the LIVT group is very likely for the same reason. This finding supports our hypotheses that there are noticeable differences between the EFL learners with different language learning backgrounds. Although this finding does not show how the different English input (from the

environment (LIVT) or from the education (TESOL)) may influence people's language use in terms of the use of words/grammar, it does show that these groups behaved differently in their discourse behaviour.

LIVT - NTE					LIVT - TE					TE - NTE				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
YEAH I	***4.060***	***0.044***	*****	+						WHEN I	7.148	0.008	**	+
Less					Less					Less				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
EH I	28.690	0.000	***	-	*AND I*	11.385	0.001	***	-	**EH I**	24.228	0.000	***	-
I I	23.167	0.000	***	-						**I I**	20.331	0.000	***	-
AND I	9.075	0.003	**	-						BECAUSE I	3.936	0.047	*	-
YES I	***8.312***	***0.004***	*****	-										
BECAUSE I	8.153	0.004	**	-										

Table 4.6 Log-likelihood Ratios: The L1 collocates of *I* in LIVT, TESOL and NTESOL

In Table 4.6, it shows that the NTESOL participants used the connective *because* in conjunction with *I* significantly more often than those in the LIVT and TESOL groups. The LIVT and TESOL participants showed no significant difference in the frequency of the use of *because I*. In fact, when we check the total use of the connective *because* in the three groups, it is found that the NTESOL participants used the connective *because* significantly more frequently than the LIVT and TESOL participants. (See Table 4.7.)

		Freq.	Per 1000 words		LLR	Sig.		
because	LIVT	311	10.77	LIVT-TE	0.13	0.719		-
	TE	329	11.08	**LIVT-NTE**	**7.57**	**0.006**	****	**-**
	NTE	313	13.43	**TE-NTE**	**5.90**	**0.015**	*****	**-**

Table 4.7 Frequencies and Log-likelihood Ratios: *because* in LIVT, TESOL and NTESOL

All the participants were asked the same questions in the interviews, and some of these questions required the participants to give reasons for their opinions. The finding that the NTESOL participants used the connective *because* more often than the LIVT and TESOL participants suggests that the NTESOL participants were prone to use the connective *because* in the context in which they constructed their reasoning in English. The LIVT and TESOL participants did use the connective

because for constructing their reasoning, but they were not prone as strongly as the NTESOL participants were.

This finding is potentially compatible with what we had discussed in chapter 2. If we consider the English input to which the NTESOL participants were exposed for learning how to reason in English (most likely from EFL coursebooks), the chance is great that *because* was the item most likely to be presented and practiced. Therefore, when the NTESOL participants faced a *why*-question, they were primed to answer with the connective *because* strongly. This situation is similar to what had happened in my previous English learning experience. When I was a freshman at the first day of our conversation course at college for my TESOL training, our lecturer greeted everyone who just got into the classroom with the sentence *'How are you?'*, and most of us replied with the sentence *'I'm fine, thank you.'* without too much hesitation. After hearing many times of *'I'm fine, thank you,'* we all laughed. Our lecturer told us that he was amazed how the Taiwanese students knew so little about replying to greetings. We were primed so strongly to use this expression, because it was the first greeting and answer that we learned from English coursebooks. When we learned various ways of replying to greeting afterwards, we used the expression *'I'm fine, thank you.'* less often. The tendency in the NTESOL group to use the connective *because* more often for giving a reason may have a similar explanation. The lack of extra English input except English coursebooks may result in the NTESOL participants being primed to associate the use of *because* strongly with a *why*-question. The LIVT and TESOL participants, instead, were more likely to have been exposed to various signals of reason. It is likely that the LIVT and TESOL participants' primings for the use of *because* to give a reason had become less strong than those of the NTESOL participants. In addition, the LIVT participants and TESOL participants were the ones who had more opportunites to use English for reasoning in their daily lives or in studies instead of using this for simple practice in language classrooms. In chapter 2, we learned from Krashen's Input Hypothesis (1982, 1985) and Swain's Output Hypothsis (1985, 1995, 2005) that comprehensible input and meaningful language practice (output) are two crucial factors in regard to successful language learning. What we have observed in the different use of *because* in these groups shows the influence from these two factors is possible.

It is also found that the TESOL participants used the collocation *when I* significantly more frequently than the NTESOL participants. There are no significant differences in the frequencies of the use of *when I* between the LIVT and TESOL groups and between the LIVT and NTESOL groups. Furthermore, it is found that the use of *when I* does not differ very much in terms of its collocates. The participants in the three groups used the patterns *when I was in junior/senior high school* and *when I was young/little* as their primary use of this collocation *when I*. The frequent use of these two

patterns is possibly relevant to what we had discussed about in our interviews. We find that there is a noticeable difference between the TESOL and NTESOL groups. However, there is no sign that suggests this difference is caused by the participants' different English learning backgrounds. The significantly greater use of *when I* in the TESOL group indicates that the TESOL participants tended to mention what had happened in the past more often than the NTESOL participants. It is an interesting tendency if we consider the fact that these TESOL participants were interviewed separately. They could not influence each other as a team. (and so couldn't the LIVT and NTESOL participants.) It makes us to wonder why a group of participants who shared the similar language learning/education backgrounds would have a similar preference of using certain kinds of discourse use/linguistic use, and why this tendency would be significantly different from another group of participants with another kind of backgrounds. The explanations to these questions may lie on how the participants were influenced by the input to which they were exposed and the life experience which made them who they were. In our hypotheses, we only focus on the factors of different input and education/training for a particular subject. However, in our finding that the TESOL participants tended to mention what had happened in the past more often than the NTESOL participants suggests there were other factors influencing people's discourse preference. We probably cannot find out what kinds of factors they are in the current study. It will need another study to investigate this in depth. Nevertheless, what we have shown is that these groups behaved differently in their language use, and this finding partly supports for our hypotheses.

Another use of the connective *and* in conjunction with *I* was used significantly more often by the TESOL and NTESOL participants in comparison with the LIVT participants. The use of the discourse markers *yeah* and *yes* in conjunction with *I* showed some noticeable differences between groups. In the following sections, we will have a closer look at these collocations.

YEAH I/ YES I

In Table 4.6, it is found that the LIVT participants used the collocation *yeah I* more often than the NTESOL participants, and the NTESOL participants used the collocation *yes I* more often than the LIVT participants. There are no significant differences between the LIVT and TESOL groups in the frequencies of the collocations *yes I* and *yeah* I, and so between the TESOL and NTESOL groups. (See Table 4.8.)

The word *yeah* is a more casual form of the affirmative discourse marker *yes*. When investigating the use of *yeah* and *yes* in general, it is found that the LIVT participants tended to use

yeah in their spoken English more than *yes* proportionally. The use of *yeah* accounts for 80.53% of the total use of *yeah* and *yes*, and the use of *yes* accounts for 19.47% of the total use of *yeah* and *yes*. Such a strong tendency of using one in preference to the other is not found in the TESOL and NTESOL groups. The use of *yeah* in the TESOL and NTESOL groups accounts for about 62% of the total use of *yeah* and *yes*, and the use of *yes* accounts for about 38% of the total use of *yeah* and *yes*. In fact, the use of *yeah* is significantly more frequent in the LIVT group in comparison with in the TESOL and NTESOL groups. The use of *yeah* is 23.07 times per 1000 words in the LIVT group. The word *yeah* is used below 20 times per 1000 words in the TESOL and NTESOL groups. The use of *yes* is significantly less frequent in the LIVT group in comparison with the use in the TESOL and NTESOL groups. The use of *yes* is 5.58 times per 1000 words in the LIVT group, 7.44 times per 1000 words in the TESOL group and 11.16 times per 1000 words in the NTESOL group. (See Table 4.9.)

	LIVT			TE			NTE		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
YES/YEAH	827	100	28.65	591	100	19.91	690	100	29.61
YES	161	19.47	**5.58**	221	37.39	**7.44**	260	37.68	**11.16**
YEAH	666	**80.53**	**23.07**	370	**62.61**	**12.46**	430	**62.32**	**18.45**
YES I	23	--	0.80	35	--	1.18	39	--	1.67
YEAH I	47	--	1.63	32	--	1.08	23	--	0.99

Table 4.8 Use of *yeah* and *yes* in LIVT, TESOL and NTESOL (1)

	LIVT - NTE				**LIVT - TE**				**TE - NTE**			
	LLR	Sig.			LLR	Sig.			LLR	Sig.		
YES/YEAH	0.41	0.522		-	46.31	0.000	***	+	50.32	0.000	***	-
YES	**49.52**	**0.000**	*******	**-**	**7.87**	**0.005**	******	**-**	**19.63**	**0.000**	*******	**-**
YEAH	**13.22**	**0.000**	*******	**+**	**94.20**	**0.000**	*******	**+**	**30.70**	**0.000**	*******	**-**
YES I	8.31	0.004	**	-	2.18	0.140		-	2.26	0.132		-
YEAH I	4.06	0.044	*	+	3.30	0.069		+	0.10	0.746		+

Table 4.9 Use of *yeah* and *yes* in LIVT, TESOL and NTESOL (2)

What we have found is that although all the three groups shows the use of the casual form *yeah* as the primary use of affirmative discourse markers. The LIVT participants behaved quite

differently from the TESOL and NTESOL participants. The LIVT participants, who had been staying in an English-speaking country and been exposed to authentic English use for at least one year, tended to use the casual form *yeah* more strongly and to use the formal form *yes* less strongly than the TESOL and NTESOL participants. This finding supports our first hypothesis that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan and vice versa. This finding shows the influence from language input, and it is potentially compatible with what has been discussed in chatper 2. The use of *yeah* is very common in spoken English. In BASEah corpus, the word *yeah* is the 70th most frequent word (825 occurrences), and the word *yes* is the 145th most frequent word (371 occurrences.) In COCAsp corpus, the word *yeah* is the 101st most frequent word (27077 occurrences), and the word *yes* is the 108th most frequent word (24191 occurrences.) The LIVT participants, who had more opportunities to hear and use spoken English, had more opportunities to be primed to use the discourse marker *yeah* than the TESOL and NTESOL participants. The Taiwanese participants from the other two groups, especially the NTESOL participants, had had fewer opportunities to hear spoken English which contains frequent use of *yeah* and fewer opportunities to put it into practice. According to Swain's Output hypothesis, the NTESOL participants therefore were unlikely to acquire this use as the LIVT participants did.

AND I

In Table 4.6, it is found that the LIVT participants used the collocation *and I* less frequently than the TESOL and NTESOL participants. In fact, the LIVT participants used the connective *and* significantly less frequently than the TESOL and NTESOL participants, and there is also no significant difference in the frequencies of the connective *and* between TESOL and NTESOL groups (See Table 4.10) The significantly more use of the connective *and* and the collocation *and I* in the TESOL and NTESOL groups or the significantly less use of *and* and *and I* in the LIVT group suggests that there are some noteworthy differences in their language use.

When investigating how these three groups of participants used the collocation *and I*, one noteworthy tendency is found in the TESOL and the NTESOL groups. (See Table 4.11.) The discourse markers *yes* and *yeah* frequently appear as the L1 collocates of *and I* in the TESOL and NTESOL groups. There is a total of 15 occurrences of the collocations *yeah and I* and *yes and I* found in the TESOL group, which accounts for 13% of their total usage of *and I*. There is a total of 13 occurrences of the collocations *yeah and I* and *yes and I* found in the NTESOL group, which accounts for 15% of their total usage of *and I* in the NTESOL group. The use of these two collocations is not found to be used as strongly as in the LIVT group. There are merely 4 occurrences of the collocations *yeah and I* and *yes and I* found in the LIVT group, which accounts for 6% of their total use of *and I*. If

we view the standardised figures, the differences between groups are much clearer. The use of collocations *yeah and I* and *yes and I* occurs only about 0.14 times per 1000 words in the LIVT group, but it occurs 0.5 times per 1000 words in the TESOL group and 0.56 times per 1000 words in the NTESOL group. The TESOL and NTESOL participants did not differ a lot in terms of the use of *yeah and I* and *yes and I*. It is the LIVT participants that made use of *yeah and I* and *yes and I* nearly 4 times less frequently than the TESOL and NTESOL participants. It is statistically significantly less frequent as shown in Table 4.12.

and I	LLR	Sig.		
LIVT – TE	11.385	**0.001**	***	-
LIVT – NTE	9.075	**0.003**	**	-
TE – NTE	0.029	0.864		+
And	LLR	Sig.		
LIVT – TE	31.03	**0.000**	***	-
LIVT – NTE	21.85	**0.000**	***	-
TE – NTE	0.33	0.567		+

Table 4.10 Log-likelihood Ratios: *and I* and *and*

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
AND I	71	100	2.46	120	100	4.04	92	100	3.95
YEAH AND I	2	3	0.07	12	10	0.40	6	7	0.26
YES AND I	2	3	0.07	3	3	0.10	7	8	0.30
YEAH/YES AND I	4	**6**	**0.14**	15	**13**	**0.50**	13	**15**	**0.56**

Table 4.11 Use of *and I* in LIVT, TESOL and NTESOL

yeah/yes and	LLR	Sig.		
LIVT - TE	0.30	0.583		-
LIVT - NTE	3.90	0.048	*	-
TE - NTE	0.07	0.794		-
yeah/yes and I	LLR	Sig.		
LIVT - TE	6.48	0.011	*	-
LIVT - NTE	7.14	0.008	**	-
TE - NTE	5.87	0.015	*	-

Table 4.12 Log-likelihood Ratios: *yeah/yes and I* and *yeah/yes and*

When examining the individual instances of these collocations, it is found that the discourse markers *yes* and *yeah* are usually used in front of the connective *and* for two purposes. One is to respond to the interviewer's opinions and then add a further clarification. It often occurs at the initial position of the utterances, as the instance TE4-1 below shows:

TE4-1 Sex_And_City &haha . # **yes and I** can understand the the I c

The other situation is that the discourse marker is used more in a sense of self-comforting or self-encouragement. It is not a response to the interviewer, but more likely to agree with himself/herself as the instances TE4-2 and TE4-3 show:

TE4-2 I have to do by myself **yes and I** think that with the help o

TE4-3 s pretty important &eh **yeah and I** plan to study abroad so En

The latter type of use is found more in the TESOL and NTESOL groups, especially when these participants were talking about themselves with the collocation *and I*. The LIVT participants used this usage less frequently in comparison with the TESOL and NTESOL participants. Moreover, when checking the reference corpora, BASEah and COCAsp, it is found that the collocations *yeah and*, *yes and*, *yeah and I* and *yes and I* are all rare. When looking at the standardised figures, the collocations *yeah and I* and *yes and I* are rarely used by the speakers in the BASEah and COCAsp corpora. (See Table 4.13.) We are not saying that the speakers in the BASEah or COCAsp corpora do not use the discourse markers as a means to assert their confidence in speaking. It is rather that the speakers in the BASEah and COCAsp corpora are not prone as strongly to use the discourse markers *yeah* and *yes* with the collocations *and* and *and I* for such a purpose. This suggests our finding that the TESOL and NTESOL participants used *and* and *and I* with *yes/yeah* so frequently is a special tendency which differentiates them from the LIVT participants and also the native speaker corpora.

	COCAsp			BASEah		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
AND	463771	100	25.13	12132	100	28.71
YEAH AND	541	0.12	0.03	28	0.23	0.07
YES AND	654	0.14	0.04	10	0.08	0.02
AND I	35364	7.63 (100)	1.92	344	2.84 (100)	0.81
YEAH AND I	70	0.02 (0.20)	0.00	2	0.02 (0.58)	0.00
YES AND I	101	0.02 (0.29)	0.01	1	0.01 (0.29)	0.00

Table 4.13 Use of *and* and *and I* in BASEah and COCAsp

The finding of significant differences between the LIVT group and the TESOL and NTESOL groups in their use of either *yeah and I* or *yes and I* partly supports the hypothesis that the EFL learners wiho go to the UK for study wll have noticeable differences in their English use in comparison with the ones in Taiwan. What we have found suggests that the TESOL and NTESOL participants used *yeah/yes and I* more often in a sense of self-comforting or self-encouragement. This is a special discourse behaviour which is rarely found in the reference corpora. The TESOL and NTESOL participants were prone to use *yeah/yes* for such a purpose more strongly than the LIVT participants were. However, it is difficult to know exactly why they were prone to use *yeah/yes* so strongly in this way in the current study. In English, we can also find some discourse markers, such as *right*, *yeah*, or *un-huh*, that people use for self-asserting. The explanation for this situation of relying on the use of *yeah*/*yes* to self-assert in the TESOL and NTESOL groups may be that there are not many discourse markers taught in EFL courses in Taiwan. The discourse markers *yeah*/*yes* are the ones most frequently taught and used in the classroom. The Taiwanese EFL students, therefore, are possible to rely heavily on them. The LIVT participants were more likely to have the opportunities to acquire other expressions to fulfil a similar discourse purpose, so they might rely less heavily on the discourse markers *yeah* or *yes* than the TESOL and NTESOL participants. This is potentially relevant to the influence on language use from language immersion discussed in chapther 2. The authentic and high-frequent input/output is what causes language learning to be achieved. The heavy use of such expressions in the TESOL and NTESOL groups also possibly indicates that the TESOL and NTESOL participants were less confident than the LIVT participants. Therefore, they displayed more self-comforting/self-asserting behaviour when speaking in English. Either way, the differences between the LIVT participants and the TESOL and NTESOL participants supports our hypotheses that there are noticeable differences between the EFL learners who go to the UK for study in comparison with the ones in Taiwan and vice versa.

4.2.2 R1 Collocates of I

Turning now to the R1 collocates of *I*, it is seen that the majority of the top 10 most frequent R1 collocates are, as before, the same in the three groups. (See Table 4.14.) The filler sound *eh* and the repetition of *I* are again found in the list, but the majority of the R1 collocates in the list are verbs: common verbs (*think, like, have* and *want*), modal verbs (*do(n't), can('t), will* and *would*) and auxiliary BE-verbs (*was*). It is worth noting that the past tense BE-verb *was* is only found in the LIVT and TESOL groups' top 10 most frequent word lists, not in the NTESOL group's. A similar situation is found in the use of adverb *just*. The R1 collocate *just* is only found in the LIVT and TESOL groups' top 10 most frequent lists, but not in the NTESOL group's list. There is also a modal verb *would* which only appears in the LIVT group's top 10 list, but not in the other two groups' lists. The lack of a similar level of high frequency for these collocates in either the TESOL or NTESOL groups suggests that there are differences between the LIVT group and the other two groups. Next, we will take a closer look at how these differences manifest themselves.

	LIVT	1826	63.25	**TESOL**	1915	64.51	**NTESOL**	1657	71.10
N	R1	Freq.	per 1000wds	R1	Freq.	per 1000wds	R1	Freq.	per 1000wds
1	THINK	507	17.56	THINK	425	14.32	THINK	354	15.19
2	DON'T	157	5.44	I	148	4.99	I	190	8.15
3	I	138	4.78	DON'T	105	3.54	DON'T	109	4.68
4	CAN	71	2.46	CAN	104	3.50	HAVE	101	4.33
5	**WAS**	69	2.39	**WAS**	98	3.30	CAN	92	3.95
6	**JUST**	67	2.32	HAVE	86	2.90	LIKE	65	2.79
7	HAVE	63	2.18	WILL	82	2.76	WILL	56	2.40
8	WILL	42	1.45	LIKE	73	2.46	EH	55	2.36
9	WOULD	37	1.28	**JUST**	36	1.21	WANT	39	1.67
10	LIKE	34	1.18	WANT	34	1.15	CAN'T	39	1.67

Table 4.14 R1 collocates of *I* in LIVT, TESOL and NTESOL

Many collocations show significant differences in the occurrences between groups. (See Table 4.15.) The LIVT participants used the collocations *I would, I just* and *I think* significantly more frequently than the TESOL and NTESOL participants. We will investigate these three collocations *I would, I just* and *I think* in the later section for examining our first hypothesis. There are other noticeable differences between the LIVT group and TESOL and NTESOL groups. The LIVT participants used the collocations *I will, I want* and *I like* significantly less frequently than the TESOL

and NTESOL participants. The significantly greater use of the collocations *I want* and *I like* in the TESOL and NTESOL groups suggests that the TESOL and NTESOL participants told about their preferences and desires more often than the LIVT participants. The collocation *I was* is used significantly more frequently in the LIVT and TESOL groups than in the NTESOL group. The collocations *I can't* and *I don't* are also found to be used in one group more than in the others. Although it will be interesting to investigate all these features, it is impossible to cover so many aspects in this study. Therefore, we selected the collocations *I was, I would, I will, I think and I just* from this list for our examination of the hypotheses.

LIVT-NTE					LIVT-TE					TE-NTE				
Over					Over					Over				
Words	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
I WOULD	**36.156**	**0.000**	*******	**+**	**I WOULD**	**19.081**	**0.000**	*******	**+**	**I WAS**	**28.934**	**0.000**	*******	**+**
I WAS	**12.068**	**0.001**	*******	**+**	I DON'T	11.888	0.001	***	+					
I JUST	**6.952**	**0.008**	******	**+**	**I JUST**	**10.359**	**0.001**	******	**+**					
I THINK	**4.422**	**0.035**	*****	**+**	**I THINK**	**9.687**	**0.002**	******	**+**					
Less					Less					Less				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
I I	23.167	0.000	***	-	I LIKE	13.483	0.000	***	-	I EH	23.721	0.000	***	-
I EH	22.559	0.000	***	-	**I WILL**	**12.047**	**0.001**	*******	**-**	I I	20.331	0.000	***	-
I HAVE	18.899	0.000	***	-	I CAN	5.375	0.020	*	-	I HAVE	7.565	0.006	**	-
I LIKE	17.643	0.000	***	-	I WANT	4.569	0.033	*	-	I CAN'T	4.363	0.037	*	-
I WANT	13.071	0.000	***	-	**I WAS**	**4.287**	**0.038**	*****	**-**	I DON'T	4.168	0.041	*	-
I CAN	9.075	0.003	**	-										
I WILL	**6.126**	**0.013**	*****	**-**										
I CAN'T	3.894	0.048	*	-										

Table 4.15 Log-likelihood Ratios: The R1 collocates of *I* in LIVT, TESOL and NTESOL

I WAS

The use of *I was* in the LIVT group is significantly less frequent than in the TESOL group, but significantly more frequent than in the NTESOL group. Interestingly, despite the fact that the overall use of the past tense BE-verb *was* in the LIVT and TESOL groups does not differ greatly, (see Table 4.16), the use of *I was* in the LIVT group is not found as many as in the TESOL group. On the other hand, the NTESOL participants did not use the past tense BE-verb *was* anything like as frequently as the LIVT and TESOL participants. (*WAS*: LIVT [115 occurrences]; TESOL [148 occurrences]; NTESOL [44 occurrences]) Unsurprisingly, therefore, the use of *I was* in the NTESOL group is significantly less frequent than in the other two groups.

What we have found here suggests that the NTESOL participants did not use past tense verbs in their spoken English as often as the LIVT and TESOL participants, even though every participant was asked to share their previous experiences in the interviews. The next step is therefore to investigate this possibility. In the LIVT and TESOL groups, the past tense BE-verbs (*was, were*) are together found to be used over 130 times in each group, whereas there are only 45 instances found in the NTESOL group. (The use of past tense BE-verbs is 4.78 times per 1000 words in the LIVT group, 5.39 times per 1000 words in the TESOL group and only 1.93 times per 1000 words in the NTESOL group.) This is true of more than past tense BE-verbs. The NTESOL participants made use of fewer past tense verbs in their spoken English. The use of past tense verbs (regular and irregular) is 66 occurrences (2.28 times per 1000 words) in the LIVT group, 118 occurrences (3.98 times per 1000 words) in the TESOL group, and merely 41 occurrences (1.76 times per 1000 words) in the NTESOL group. (See Table 4.17.) All of these findings suggest that the NTESOL participants behaved very differently from the LIVT and TESOL participants in terms of the use of past tense.

I WAS	LLR	Sig.		
LIVT - TE	4.287	**0.038**	*	-
LIVT - NTE	12.068	**0.001**	***	+
TE - NTE	28.934	**0.000**	***	+
WAS	LLR	Sig.		
LIVT - TE	3.28	0.070		-
LIVT - NTE	19.47	**0.000**	***	+
TE - NTE	37.11	**0.000**	***	+

Table 4.16 Log-Likelihood Ratios: *I was* and *was*

	LIVT		**TE**		**NTE**	
(Past Particles Excluded)	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
Regular Past Tense Verbs	20	0.69	62	2.09	13	0.56
Irregular Past Tense Verbs	46	1.59	56	1.89	28	1.20
Past Tense *DO*-Verb (DID)	42	1.45	70	2.36	26	1.12
Past Tense *BE-Verb*s	**138**	4.78	**160**	5.39	**45**	1.93

Table 4.17 Use of Past Tense Verbs in LIVT, TESOL and NTESOL

The finding that the NTESOL group differs from the LIVT and TESOL groups in its lack of use of past tense partly supports the hypothesis that the EFL learners who go to the UK for study and

the EFL learners who major in English-relevant subjects will have noticeable differences in their English use in comparison with the ones who do not have these backgrounds. The participants in the NTESOL group were non-English major EFL students. They had fewer opportunities to use English in their daily lives and also in their studies. That is to say, where they had had opportunities to use the past tense, they were more likely to have occurred in written English (for examinations or assignments), not in spoken English. Of course they were taught how to use the past tense in the EFL classroom, and they understood the use of past tense in English. However, the low frequency of real opportunities to hear and speak might have resulted in their being weakly primed in the use of these past tense verbs in their spoken English. Therefore, it can possibly explain why the NTESOL participants described things in the present tense when they actually told about the past events. As the instance shows in Table 4.18, when the participant NTE07 mentioned her friends' previous working experiences, she only used one past tense verb *went*, and the rest verbs were in the present tense. It shows that the participant NTE07 had processed the knowledge to use the past tense initially but dropped it afterwards. The finding is potentially compatible with what has been discussed in chapter 2. Learners who have receptive vocabulary knowledge (know and recognise the meaning of words) may not be able to use this knowledge productively (Nation, 2001; Webb, 2005, 2008). Our finding suggests that NTE07 might have been primed to use the past tense forms of the words receptively but not productively. Since NTE07 was a EFL learner who did not have to use English in her daily life and in her study, the lack of proper opportunities for her to actively produce those words might be the explanation of this situation.

&eh &eh I &eh my friend I have some friend &eh some friends go &eh **went** to work last year and I &eh I think they I think it's &eh the pay **is** too low but they **have** to do too much works and &eh and can't (..) and can't &eh too much but they **don't** get the equal pay yeah and even if they **get** much more higher pay they still **have** to do more more job yeah .

Table 4.18 Instance of the Verbs used by NTE07

I WOULD

In the use of the collocation *I would*, a special tendency in the LIVT group is found. The use of *I would* in the LIVT group is significantly more frequent than in the TESOL and NTESOL groups. In fact, this difference does not only occur in the use of the collocation *I would*; the total use of the modal verb *would* in the LIVT group is also more frequent than in the other two groups. (See Table 4.19.) There are 70 occurrences of *would* found in the LIVT group (2.42 times per 1000 words). There

are only 18 occurrences of *would* in the TESOL group (0.61 times per 1000 words) and 3 occurrences in the NTESOL group (0.13 times per 1000 words.) (See Table 4.20.)

I WOULD	LLR	Sig.		
LIVT - TE	19.081	**0.000**	***	+
LIVT - NTE	36.156	**0.000**	***	+
TE - NTE	5.57	**0.018**	*	+
WOULD	LLR	Sig.		
LIVT - TE	34.29	**0.000**	***	+
LIVT - NTE	62.66	**0.000**	***	+
TE - NTE	8.56	**0.003**	**	+

Table 4.19 Log-Likelihood Ratios: *I would* and *would*

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WOULD	70	100	2.42	18	100	0.61	3	100	0.13
I WOULD	37 (100%)	53	1.28	9(100%)	50	0.30	1(100%)	33	0.04
R1 Collocates of *I WOULD*	**LIKE 14(38%)** **SAY 13(35%)** NOT 2 JUST 2 ACTING 1 TALK 1 LEARN 1 LISTEN 1 RATHER 1 GO 1			**LIKE 4(44%)** APPLY 1 就(jiù) 1 BE 1 CHECK 1 LISTENING 1			**LIKE 1(100%)**		

Table 4.20 Use of *I would* and *would* in LIVT, TESOL and NTESOL

The finding that the LIVT group differs from the TESOL and NTESOL groups in its use of either *I would* or *would* partly supports our hypothesis that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan. In the reference corpora, it is found that the use of *would* is 2.13 times per 1000 words in the BASEah corpus and 2.47 times per 1000 words in the COCAsp corpus. The frequency of the use of *would* in the LIVT group is very similar to the one in BASEah and COCAsp. The use of *would* in the LIVT group is 2.42 times per 1000 words. It is also found that the words *like* and *say* are common R1

collocates of *would* (*would like/would say*) in BASEah and COCAsp. (See Table 4.21.) Although the BASEah and COCAsp corpora are not the exact input to which the LIVT participants encountered, the finding suggests that the LIVT participants had a similar behaviour in the use of the word *would* in comparison with the native speakers. What we see here is compatible with what has been discussed in chapter 2. As Kemmer and Israel (1994: 167) state, "the more speakers talk to each other the more they will talk alike, and so linguistic variation will pattern along lines of social contact and interaction." The LIVT participants were the ones who had the most opportunities to talk to native speakers of all the participants. The finding supports that they were very likely influenced by what they had encountered in their language contact and therefore had been primed to use English in this way. This may be a sign of acculturation. However, whether this acculturation happened intentionally or unintentionally still warrants for more research.

	BASEah			**COCAsp**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WOULD	898	100	2.13	45532	100	2.47
I WOULD	97	10.80	0.23	7657	16.82	0.41
Patterns of *would*						

N	L1	Centre	R1
1	IT	WOULD	BE
2	THAT		HAVE
3	YOU		SAY
4	HE		YOU
5	THEY		LIKE

N	L1	Centre	R1
1	IT	WOULD	BE
2	THAT		HAVE
3	HE		N'T
4	YOU		YOU
5	THEY		SAY

Table 4.21 Use of *I would* and *would* in BASEah and COCAsp

The lack of use of *would* and *I would* in the TESOL and NTESOL groups is also potentially compatible with what has been discussed. The TESOL and NTESOL participants, who were located in Taiwan and had fewer opportunities to be exposed to authentic English input and to actually use English in daily lives, seemed not yet to be primed, or were only weakly primed, to use the modal verb *would*. Furthermore, it can be seen that when these Taiwanese students did use the modal verb *would*, they were strongly prone to use this word with the pronoun *I*. Of the total instances of the use of *would* in the LIVT and TESOL groups, half (LIVT: 53%; TE: 50%) are found to be with the first person pronoun *I*. Although this tendency is not so pronounced in the NTESOL group, it is still the case that a third of the instances of *would* occur in the combination *I would*. When checking the R1 collocates of *I would*, it is found that both the TESOL and NTESOL participants were strongly prone to use the pattern *I would like (to)*. The use of the pattern *I would like (to)* is explicitly taught in

English learning materials, therefore, it may suggest that the TESOL and NTESOL participants might be influenced by their learning materials. Although the LIVT participants were also prone to use the pattern *I would like (to)*, they were prone to use another pattern *I would say* in a similar proportion. (*I would like*: 38%; *I would say*: 35%). These findings suggest that the use of the word *would* in these three groups is different in terms of the word's collocations. The LIVT participants used *would* much more often and more flexibly than the TESOL and NTESOL participants, and the different langaugae environments seem to be the reason for this. Next, we will investigate the use of *I WILL*.

I WILL

The LIVT participants used the collocation *I will* significantly less frequently than the TESOL participants, and slightly less frequently than the NTESOL participants. (See Table 4.22.)

I will	LLR	Sig.		
LIVT - TE	12.047	**0.001**	***	-
LIVT - NTE	6.126	**0.013**	*	-
TE - NTE	0.652	0.419		+
will	LLR	Sig.		
LIVT - TE	5.68	**0.017**	*	+
LIVT - NTE	13.18	**0.000**	***	+
TE - NTE	2.01	0.156		+

Table 4.22 Log-Likelihood Ratios: *will* and *I will*

However, when examining the overall use of the word *will*, it is found that the LIVT participants actually used the modal verb *will* more frequently than the TESOL and NTESOL participants. There are 209 occurrences of the word *will* in the LIVT group, while there are 168 occurrences in the TESOL group and 111 occurrences in the NTESOL group. Nearly fifty percent of the instances of *will* in the TESOL and NTESOL groups is the collocation *I will* (TESOL: 49%; NTESOL: 50%). However, in the LIVT group, the use of *I will* merely accounts for 20% of the total use of *will*. The LIVT participants tended to use *will* with the second person pronoun *you* more than with the first person pronoun *I*. There are 49 occurrences of *you will* and 42 occurrences of *I will* in the LIVT group. Such a distribution is not found in the TESOL and NTESOL groups. (See Table 4.23.)

These finding suggests that the LIVT participants made more use of the modal verb *will* in their spoken language, and the way they used the modal verb *will* was different from the TESOL and NTESOL participants. When the TESOL and NTESOL participants used *will* in their speech, it was mostly about what they intended to do or what they were capable of doing. Such a usage was also employed by the LIVT participants, but the LIVT participants were prone to use the modal verb *will* with *you* or *they* in a general sense, that is, to state a potential situation or a general fact, as shown in Table 4.24:

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WILL	209	100	7.24	168	100	5.66	111	100	4.76
I WILL	42	**20**	1.45	82	**49**	2.76	56	**50**	2.40
Other L1 collocates of *WILL*	**YOU 49** THEY 29 IT 27 THAT 14 WE 5			THEY 25 IT 11 TEACHER 7 **YOU 6** WE 5 PEOPLE 5			EH 23 **YOU 7** IT 5 WE 5		

Table 4.23 Use of *will* and *I will* in LIVT, TESOL and NTESOL

N	Concordance
7	yeah so I don't think &eh I will feel a little bit pressure
8	hen I speaking in English I will be more nervous than speaki
9	&eh if I like the teacher I will &eh like the language and d
10	school to study English I I will not feel not so uncomfortab
11	tise so I (...) I don't I I will not I will not hate cram sc

Instances from NTESOL

N	Concordance
161	yeah but but sometimes you will only focus on subtitle not
162	you study in Liverpool you will &t you you will listen the
163	you to use language or you will be more more fluent , &eh ,
164	verpool you will &t you you will listen the scouser accent i
165	(..) I think it will be you will feel like you have a conver
166	not catch the speed and you will lose this part of meaning a

Instances from LIVT

Table 4.24 Instances of *will* used in LIVT and NTESOL

The differences in the use of *will* between the LIVT group and the TESOL and NTESOL groups partly support the hypothesis that the EFL learners who go to the UK for study will have noticeable

differences in their English use in comparison with the ones in Taiwan. However, these differences seem not to be influenced by the English input they received from their different language learning environments, but more likely to be the results of different discourse preferences. The LIVT participants focused less on what they intended to do or were capable of doing. They focused more on the potential situation or general facts.

I JUST/ I THINK

In the previous section, it is found that the collocations *I just* and *I think* were employed significantly frequently by the LIVT participants in comparison with the TESOL and NTESOL participants. When checking the total use of the words *just* and *think*, it is found that the LIVT participants also used these two words *just* and *think* significantly more frequently than the TESOL and NTESOL participants. (See Table 4.25 and Table 4.26.)

I just	LLR	Sig.		
LIVT - TE	10.359	**0.001**	**	+
LIVT - NTE	6.952	**0.008**	**	+
TE - NTE	0.142	0.706		-
just	LLR	Sig.		
LIVT - TE	54.61	**0.000**	***	+
LIVT - NTE	10.34	**0.001**	**	+
TE - NTE	14.21	**0.000**	***	-

Table 4.25 Log-Likelihood Ratios: *I just* and *just*

I think	LLR	Sig.		
LIVT - TE	9.687	**0.002**	**	+
LIVT - NTE	4.422	**0.035**	*	+
TE - NTE	0.675	0.411		-
think	LLR	Sig.		
LIVT - TE	17.20	**0.000**	***	+
LIVT - NTE	7.47	**0.006**	**	+
TE - NTE	1.36	0.244		-

Table 4.26 Log-Likelihood Ratios: *I think* and *think*

In English, the word *just* is used as an adverb mostly, which is a means for speakers to soften the tone of a statement (a) or to assign to the statement a limited degree of certainty (b). As the example shows:

(a)	ctise my listening and now I just think it's okay now I can
(b)	g without the subtitle and I just guess (..) their what are

The word *think* is a verb used to state that the speaker has a thought or believes that something is true. However, the collocation *I think* occurs so often that it seems to be a habitual starting interjection when a speaker starts a sentence, as shown in Table 4.27:

N	Concordance
1	of accent but (...) but I I think I already practise a lot t
2	start from this year yeah I think first I watch the film wit
3	tch film without subtitle I think it's start from this year
4	e movie afterward but but I think after you first time you w

Table 4.27 Instances of *I think* used in LIVT

It carries little significance and it is strictly redundant, since what a speaker says is often what s/he thinks about. The use of *I think* at the beginning of sentences might be viewed as a discourse strategy, such as presenting a certain level of modesty and also gaining more time for the speaker to articulate what s/he has to say. The frequent use of *I just* and *I think* in the LIVT group suggests that the LIVT participants tended to make use of hedging more often than the TESOL and NTESOL participants in their spoken English. The findings partly support the hypothesis that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan. The use of the words *(I) just* and *I think* in such a discourse function is commonly found in authentic English. The LIVT participants were exposed to this kind of language use more frequently than the TESOL and NTESOL participants were. The finding that the LIVT participants used these two collocations much more often than the TESOL and NTESOL participants suggests that the LIVT participants might be influenced by their English input greatly. What we see here is again compatible with what has been discussed in chapter 2, that is, the more speakers talk to each other the more they will talk alike. The input, either from English learning materials or from authentic English input of the environment, influences how people are primed to use words. When investigating the collocates of *think*, it is found that the LIVT and TESOL participants tended to use the word *that* significantly more often than the NTESOL participants, and the TESOL participants used the word *that* significantly more often than the LIVT participants. (See Table 4.28.)

think that	LLR	Sig.		
LIVT – TE	14.476	**0.000**	***	-
LIVT – NTE	20.626	**0.000**	***	+
TE – NTE	59.532	**0.000**	***	+

Table 4.28 Log-likelihood Ratios: *think that*

There are 2.16 times per 1000 words of the use of the collocation *think that* in the TESOL group, which is twice as frequent as in the LIVT group (0.94 times per 1000 words) and 24 times more frequent than in the NTESOL group (0.09 times per 1000 words). (See Table 4.29.) The word *that* in the collocation *think that* is utilised to create a subordinate clause, and this usage is common in English, especially in written English or in formal English. The use of *think that* is found 2.64 times per 1000 words in the BASEah corpus and 2.53 times per 1000 words in the COCAsp corpus. (See Table 4.30.) The BASEah corpus contains academic spoken English and the COCAsp corpus contains formal English from TV news or interviews. The major input that the participants in the TESOL group had received was formal academic English (mostly in written English), and these participants usually produced English mainly for academic purposes (in the EFL classrooms and for essay writing). Therefore, the TESOL participants were likely to be influenced to use the collocation *think that* more strongly than the other participants. The participants in the LIVT group also had received and produced English mainly in the academic environment, but they were more likely to encounter casual English in their daily lives. Therefore, the LIVT participants might therefore be primed less strongly to use *think* with *that* as frequent as the TESOL participants. Another possible explanation is contextualisation that they might be influenced to use *think that* in specific and appropriate contexts. Nevertheless, the findings of how the TESOL and LIVTparticipants used the collocation *think that* partly support the hypothesis that the EFL learners who study in English-relevant subjects will show noticeable differences in their use of English from those studying in non- English-relevant subjects. The findings also partly support the hypothesis that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan.

	LIVT	559	100	19.36
N	**R1**	**Freq.**	**%**	per 1000wds
1	IT'S	123	20.53	4.26
2	EH	59	9.85	2.04
3	THE	46	7.68	1.59
4	I	42	7.01	1.45
5	IT	40	6.68	1.39
6	**THAT**	**27**	**4.51**	**0.94**
7	THEY	16	2.67	0.55
8	SO	14	2.34	0.48
9	MAYBE	13	2.17	0.45
10	ENGLISH	12	2.00	0.42
	TESOL	478	100	16.10
N	**R1**	**Freq.**	**%**	per 1000wds
1	IT'S	98	20.50	3.30
2	**THAT**	**64**	**13.39**	**2.16**
3	EH	57	11.92	1.92
4	I	31	6.49	1.04
5	THE	30	6.28	1.01
6	THAT'S	20	4.18	0.67
7	IT	16	3.35	0.54
8	ENGLISH	11	2.30	0.37
9	THEY	8	1.67	0.27
10	SO	8	1.67	0.27
	NTESOL	406	100	17.42
N	**R1**	**Freq.**	**%**	per 1000wds
1	IT'S	126	31.03	5.41
2	EH	59	14.53	2.53
3	THE	34	8.37	1.46
4	I	29	7.14	1.24
5	IT	18	4.43	0.77
6	ENGLISH	14	3.45	0.60
7	THEY	12	2.96	0.51
8	THIS	7	1.72	0.30
9	IF	7	1.72	0.30
10	SO	5	1.23	0.21
--	**THAT**	**2**	**0.5**	**0.09**

Table 4.29 R1 collocates of *think* in LIVT, TESOL and NTESOL

	BASEah			COCAsp		
Words	Freq.	%	per 1000 words	Freq.	%	per 1000 words
think	1,876	100	4.44	72,575	100	3.93
think that	1114	59.38	2.64	46723	64.38	2.53

Table 4.30 Use of *think* and *think that* in BASEah and COCAsp

4.2.3 Summary of the use of *I*

In 4.2, we have investigated the top 10 most frequent collocates of the pronoun *I* in the LIVT, TESOL and NTESOL groups. It has been found that several collocates of *I* which show noticeable differences in the frequency between the three groups. Several findings indicate that differences in language use exist between these three groups. In some cases, the causes of the differences may relate to the different environment input and pedagogical input received by the groups. Some differences relate more to the different discourse strategies preferred by one particular group. Regarding the latter one, it is difficult to directly link the cause of the differences in discourse preferences to the participants' input/background differences. However, what we have found in this kind of use still partly supports our first and second hypotheses that the EFL learners with different language learning backgrounds will behave differently in their language use. The findings in the investigation of the use of *I* are presented in Table 4.31.

We have found several differences that partly support our first and second hypotheses. In the use of the collocation *yeah I* and *yes I*, we found that the LIVT participants tended to use *yeah (I)* significantly more often than the NTESOL participants, and the NTESOL participants tended to use *yes (I)* more often than the LIVT participants. The LIVT participants, who had received abundant authentic and various input from the environment and English-speaking classrooms, preferred to use *yeah I* as their primary use, and the participants in the NTESOL group, who of the three groups had received the least quantity and least variety of input from the EFL classrooms, preferred to use *yes I* as their primary use. The TESOL participants, who had received not so much authentic input from the environment but had had rich pedagogical input from the classrooms, used both *yeah I* and *yes I* in a similar proportion. The use of the collocation *yeah/yes and I* also reveals how the groups of participants behaved differently in their language use. The TESOL and NTESOL participants preferred to use the *yes*/*yeah* with *and I* with the apparent purpose of self-comforting/asserting. The LIVT participants, however, did not possess this preference as strongly as the TESOL and NTESOL participants. The LIVT participants, in terms of using the discourse markers *yes*/*yeah* as collocates of

I, were more similar to the way the speakers in the BASEah and COCAsp corpora did. The findings with regard to the use of *yeah/yes (and I)* support our hypotheses that the EFL learners who go to the UK for study will differ from the ones in Taiwan in their language use. The findings are, to some degree, compatible with what has been discussed in chapter 2. Comprehensible input, meaningful output along with the practice of negociation are the key elements which influence people's language learning. When the use of words occurs frequently, it is likely to become stored in a person's long-term memory and easier to be retrieved afterwards. Therefore, people are likely to be strongly primed to use the words/combination of words to which they are frequently exposed. This leads to what Kemmer and Israel (1994: 167) state, "the more speakers talk to each other the more they will talk alike, and so linguistic variation will pattern along lines of social contact and interaction." What we have observed suggests that the immersion in different language environments and the diversity of school education are very likely to cause people to be primed differently in their language use.

Support for Hypothesis 1		
Differences	Possible Causes of the differences	Words
The LIVT participants used the words **significantly more frequently** than the TESOL and NTESOL participants.	Input Influence	*yeah I/ yeah*; *I would/ would*
	Different discourse preferences	*I just/ just; I think/ think; I will/ will*
The LIVT participants used the words **significantly less frequently** than the TESOL and NTESOL participants.	Input Influence	*yes I/yes*
	Different discourse preferences	*yeah and I/yes and I*
Support for Hypothesis 2		
Differences	Possible Causes of the differences	Words
The TESOL participants used the words **significantly more frequently** than the LIVT and NTESOL participants.	Input Influence	*think that/that*

Table 4.31 Findings in the investigation of the use of *I*

In the investigation of the collocation *I was*, we found is also potentially compatible with what has been discussed in chapter 2. The finding that the NTESOL participants used the past tense infrequently may indicate how the frequency of exposure to a grammatical form (input) and the actual practice of this grammatical form (output) could lead to differences in people's language use. In comparison with the LIVT and TESOL participants, the NTESOL participants had been influenced less strongly on the use the past tense form of verbs (regular, irregular and auxiliary BE- and DO-verbs) in their spoken English. The NTESOL participants were the ones who had fewer opportunities to receive English language input and to produce English in daily lives and in their

studies. Therefore, although beyond doubt the NTESOL participants understood the past tense verbs in English, they simply did not use the past tense very often even when they were talking about past events. This finding is again compatible with what has been discussed in chapter 2. There are two kinds of knowledge: receptive knowledge and productive knowledge. Learners who have receptive vocabulary knowledge (know and recognise the meaning of words) may not be able to use this knowledge productively (Nation, 2001; Webb, 2005, 2008).

In the use of the modal verbs *will* and *would,* the TESOL and NTESOL participants were found to be different from the LIVT participants. First, the LIVT participants were more likely to utilise the two modal verbs *will* and *would* in their spoken English, particularly in collocation with *I*. Also, the TESOL and NTESOL participants tended to use the fixed pattern *I would like (to)* when they used the word *would* in their speech. Although the LIVT participants also used the pattern *I would like (to)* frequently in their spoken English, the LIVT participants appeared to have more flexible and more varied use of these words. Another finding concerned the modal verb *will*. The LIVT participants were more likely to use the modal verb *will* with *you* or *they* in a general sense (to state a potential situation or a general fact). Such a tendency was not found in the TESOL and NTESOL groups. These two findings provide further support for hypothesis 1. We also found support for hypothesis 1 in the discovery that the LIVT participants tended to utilise certain collocations for discourse purposes which the TESOL and NTESOL participants were not prone to use in their spoken English. The comparatively frequent use of *I just* and *I think* in the LIVT group is the evidence for this difference. Our findings regarding the use of *think that* in the TESOL group support hypothesis 2, that the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying in non-English- relevant subjects. The explicit pedagogical input that the TESOL participants had might have influenced them to use *think* with *that* more strongly than the other participants. In summary, the finding is also potentially compatible with what has been discussed in chapter 2. Comprehensible input, meaningful output along with the practice of negociation are the key elements which influence people's language learning. People are likely to be strongly primed to use the words/combination of words to which they are frequently exposed. In the next section, we will investigate the use of the English second person pronoun *you*.

4.3 The Use of Pronoun *you*

At the beginning of this chapter, we established that the LIVT participants seemed to have a strong tendency to employ the second person pronouns *you* and 你 [you] more often than the TESOL and NTESOL participants. This observation was confirmed when we compared the use of *you* and 你 [you] in terms of the log-likelihood ratios. In the use of the English second person pronoun *you*, the LIVT participants showed a significantly more frequent use than the other participants. (See Table 4.32.) The TESOL participants also revealed a strong tendency to use *you* more often than the NTESOL participants. More surprisingly with regard to the use of the Mandarin second person pronoun 你 [you], the LIVT participants also had a significantly higher frequency of use in comparison with the TESOL and NTESOL participants. This is surprising because it would be expected that the Chinese language behaviour of the three groups would be unaffected by their different exposure to English and they acquired Mandarin under similar circumstances. These findings with respect to 你 [you] suggest that, although the participants in all groups were asked identical questions, the LIVT participants seemed to have behaved differently from the TESOL and NTESOL participants by making greater use of *you* and 你 [you] in their answers. This finding challenges our third, null, hypothesis that there will be no differences in the use of Mandarin Chinese between the LIVT group and the Taiwan-based groups. In the following section, we will investigate the collocations of the use of English second person pronoun *you*. The Mandarin pronoun 你 [you] will be investigated in Chapter 5.

		Freq.	Per 1000 words		LLR	Sig.		
you	LIVT	714	24.73	**LIVT-TE**	58.077	**0.000**	***	+
	TESOL	299	10.07	**LIVT-NTE**	121.358	**0.000**	***	+
	NTESOL	236	10.13	**TE-NTE**	8.843	**0.003**	**	+
		Freq.	Per 1000 words		LLR	Sig.		
你 (nǐ)	LIVT	1014	26.62	**LIVT-TE**	186.871	**0.000**	***	+
	TESOL	632	18.18	**LIVT-NTE**	160.364	**0.000**	***	+
	NTESOL	637	15.38	TE-NTE	0.004	0.951		-

Table 4.32 Log-Likelihood Ratios: *you* and 你(nǐ) in LIVT, TESOL and NTESOL

4.3.1 L1 Collocates of *you*

In the lists of the top 10 most frequent L1 collocates of English pronoun *you*, the connectives *if, and,* and *when* are found in all three groups. So are the filler sound *eh* and the repetition of the pronoun *you*. The connectives *because* and *so* are however only found in the LIVT and TESOL groups' lists, and the connective *but* is only found in the TESOL group's list. There are several L1 collocates of other kinds that occur only in one group's top 10 most frequent L1 collocates list. The words *what* and *then* and the word *that* in its *that*-clause function are found in the LIVT group's list only, and the preposition *to* and the auxiliary *do* are likewise found in the TESOL group's list only. The verbs *tell, give, teach* and the modal verb *can* are found only in the NTESOL group's list. (See Table 4.33.)

you	**LIVT**	714	100	24.73	**TE**	299	100	10.07
N	L1	Freq.	%	per 1000 words	L1	Freq.	%	per 1000 words
1	IF	81	11.34	2.81	IF	24	8.03	0.81
2	AND	49	6.86	1.70	EH	19	6.35	0.64
3	EH	48	6.72	1.66	AND	18	6.02	0.61
4	YOU	39	5.46	1.35	WHEN	16	5.35	0.54
5	WHEN	29	4.06	1.00	BECAUSE	12	4.01	0.40
6	SO	24	3.36	0.83	TO	10	3.34	0.34
7	BECAUSE	17	2.38	0.59	SO	9	3.01	0.30
8	WHAT	15	2.10	0.52	YOU	8	2.68	0.27
9	THEN	14	1.96	0.48	DO	8	2.68	0.27
10	**THAT**	14	1.96	0.48	BUT	8	2.68	0.27
-					*THAT*	*0*	*0*	*0*
	NTE	236	100	10.13				
N	L1	Freq.	%	per 1000wds				
1	IF	40	16.95	1.72				
2	AND	27	11.44	1.16				
3	EH	20	8.47	0.86				
4	YOU	9	3.81	0.39				
5	WHEN	9	3.81	0.39				
6	TELL	9	3.81	0.39				
7	GIVE	7	2.97	0.30				
8	CAN	7	2.97	0.30				
9	TEACH	6	2.54	0.26				
10	DO	6	2.54	0.26				
--	*THAT*	*1*	*0.42*	*0.04*				

Table 4.33 L1 Collocates of *you* in LIVT, TESOL and NTESOL

LIVT - NTE					LIVT - TE				
OVER					OVER				
Words	LLR	Sig.			Words	LLR	Sig.		
SO YOU	17.528	0.000	***	+	**IF YOU**	**34.284**	**0.000**	*******	**+**
YOU YOU	14.340	0.000	***	+	YOU YOU	19.156	0.000	***	+
BECAUSE YOU	14.009	0.000	***	+	AND YOU	15.781	0.000	***	+
THAT YOU	10.834	0.001	***	+	**WHAT YOU**	**15.092**	**0.000**	*******	**+**
WHAT YOU	**8.663**	**0.003**	******	**+**	EH YOU	13.797	0.000	***	+
WHEN YOU	**7.228**	**0.007**	******	**+**	SO YOU	7.499	0.006	**	+
IF YOU	**6.776**	**0.009**	******	**+**	**WHEN YOU**	**4.180**	**0.041**	*****	**+**
EH YOU	6.662	0.010	**	+					
LESS					LESS				
					Words	LLR	Sig.		
					TO YOU	5.602	0.018	*	-

TE - NTE				
Over				
Words	LLR	Sig.		
BECAUSE YOU	8.499	0.004	**	+
LESS				
Words	LLR	Sig.		
IF YOU	**8.849**	**0.003**	******	**-**
TELL YOU	6.673	0.010	**	-
GIVE YOU	6.631	0.010	*	-
AND YOU	4.647	0.031	*	-

Table 4.34 Log-Likelihood Ratios: The L1 collocates of *you* in LIVT, TESOL and NTESOL

Examining the word lists and the log-likelihood ratios (see Table 4.34), it is found that the use of English pronoun *you* differ across groups. The LIVT participants used the collocation *you you*, *eh you*, *so you*, *what you*, *if* and *when you* significantly more frequently than the TESOL and NTESOL participants. The LIVT and TESOL participants used the collocation *because you* significantly more frequently than the NTESOL participants. The TESOL participants used the collocation *and you* significantly less frequently than the LIVT and NTESOL participants. The use of repetition and fillers (*you you* and *eh you*), as we had discussed in the previous section, is a common speech behaviour in

speaking. Therefore, we will not investigate these two collocations. The use of *because you* and *so you* is used for illustrating reasons. We had viewed how these participants used *because* in the previous discussion. We did not find the use of *so* as a connective had too much difference in its frequencies between groups. Therefore, we will not have a further investigation on these two collocations. The same reason applies to the use of *and you*. The significantly high frequency of the pronoun *you* and its collocations *if you*, *when you*, and *what you* in the LIVT group supports the hypothesis 1 that the EFL learners who go to the UK for study will behave differently in their language use from the ones in Taiwan. In the use of *if you*, there are also some noticeable differences in the frequencies between the TESOL and NTESOL groups. The TESOL participants used *if you* significantly less frequently than the NTESOL participants. This finding partly supports the hypothesis 2 that the EFL learners studying in English-relevant subjects will behave differently in their language use from the ones studying in non-English-relevant subjects. Next, we will examine how the L1 collocations *if you*, *when you*, and *what you* are used in the three groups.

IF YOU

The connective *if* is the most frequent L1 collocate of *you* in all three groups. The LIVT participants, in particular, used the collocation *if you* in their spoken English more often than the TESOL and NTESOL participants. There are 81 occurrences of *if you* found in the LIVT group as opposed to only 24 occurrences found in the TESOL group and 40 occurrences in the NTESOL group. In fact, it is not only the use of *if you* that differentiated the LIVT participants from the ones in the other two groups. The use of the connective *if* itself is also more frequent in the LIVT group than in the TESOL group. (See Table 4.35.)

if you	LLR	Sig.		
LIVT - TE	34.284	**0.000**	***	+
LIVT - NTE	6.776	**0.009**	**	+
TE - NTE	8.849	**0.003**	**	-
if	LLR	Sig.		
LIVT - TE	24.45	**0.000**	***	+
LIVT - NTE	1.50	0.221		+
TE - NTE	11.94	**0.001**	***	-

Table 4.35 Log-Likelihood Ratios: *if you* and *if*

The LIVT participants used the collocation *if you* more frequently than the TESOL and NTESOL participants, which provides further support for hypothesis 1. There are 81 occurrences of *if you* in the LIVT group, which accounts for 38% of the total use of *if*. The use of *if you* in the TESOL group only accounts for 19% of the total use of *if* (24 occurrences) and 26% of the total use of *if* in the NTESOL group (40 occurrences). When examining the use of the connective *if* in the TESOL and NTESOL groups, it is found that these participants, the TESOL participants in particular, were more likely to use the connective *if* with the pronoun *I* in their speech. The use of *if I* in both TESOL and NTSOL groups accounts for about 30% of the total use of *if*, while it only accounts for 21% of the total use of *if* in the LIVT group. This finding suggests that when applying the conditional *if* in the spoken English, the LIVT participants tended to use it with the second person pronoun *you* more strongly than the TESOL and NTESOL participants. The differences between groups are more distinct when we observe the standardised frequencies. The use of *if you* is 2.81 times per 1000 words in the LIVT group, 0.81 times per 1000 words in the TESOL group and 1.72 times per 1000 words in the NTESOL group. The use of *if you* in the LIVT group is three times more frequent than in the TESOL group. The NTESOL participants even used *if you* twice more frequently than the TESOL participants. (See Table 4.36.)

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
IF	213	100	7.38	127	100	4.28	151	100	6.48
IF YOU	81	38	**2.81**	24	19	**0.81**	40	26	**1.72**
IF I	45	21	1.56	38	38	1.28	44	29	1.89
Other R1 collocates of *IF*	**YOU 81 (38%)** **I 45 (21%)** THE 15 IF 15 THEY 7 WE 6 LIKE 5 IT'S 5 EH 5			**I 38 (30%)** **YOU 24 (19%)** EH 10 WE 9 IF 7 THE 6 THEY 3			**I 44 (29%)** **YOU 40 (26%)** IF 17 EH 10 THE 8 THEY 5 WE 4 IN 4 MY 3 ENGLISH 3		

Table 4.36 Frequency of use of *if* and *if you* and other collocates of *if* in LIVT, TESOL and NTESOL

When checking these results against the two reference corpora, BASEah and COCAsp, it is found that the use of *if you* is the primary collocation of *if*. There are 1737 occurrences of *if* in the BASEah corpus. The collocation *if you* is the most frequent collocation in BASEah and it accounts for

52% of the total uses of *it*. In the COCAsp corpus, the collocation *if you* is also the most frequent collocation of the use of *if*, and it accounts for 33% of the total use of *if*. (See Table 4.34.) The BASEah corpus is, as previously noted, a collection of lecture speech and the COCAsp is made up of TV news and interviews. All of these genres usually use many conditionals but, unlike the situation of the interviews in this study, they usually have little relation to personal affairs, but are used to express impersonal generalisation. Therefore, the high frequency of the use of *if you* in these corpora is expected. However, since the interviews with participants in the three groups related to their personal affairs, there is a need to explain why the LIVT participants showed a distinct tendency to use the conditional *if you* strongly in their language use in a way that the TESOL and NTESOL participants did not. Examination of the instances of *if you* in the LIVT data reveals that the LIVT participants utilised the collocation *if you* frequently for the purpose of providing a general statements or illustration. The LIVT participants were presumably exposed to authentic English-language lectures and TV news more than the TESOL and NTESOL participants, and such exposure would have provided them with abundant examples of the use of *if you* in a general sense. Hence, it is very likely that the LIVT participants were influenced accordingly. What we see here is again compatible with what has been discussed in chapter 2. The input, either from English learning materials or from authentic English input of the environment, influences how people are primed to use words.

	BASEah			**COCAsp**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
IF	1737	100	4.11	54236	100	2.94
IF YOU	904	**52**	2.14	17752	**33**	0.96
IF I	75	4	0.18	3612	7	0.20

Table 4.34 Frequency of use of *if* ,*if you* and *if I* in BASEah and COCAsp

N	Concordance
1	hers they don't care about **if you** use the correct grammar ye
2	eryone learned English and **if you** want to &eh get a good job
3	language of the world and **if you** speak English you can lear
4	onic you need language and **if you** if you &eh &eh want to get
5	for your own interest and **if you** are interesting in want to

Table 4.35 Instances of *if you* in LIVT

The LIVT participants used *if you* more frequently in their spoken English, and the way the collocation *if you* was used in the three groups was also quite different. We can see in Table 4.36 that the patterns of the collocation *if you* differ between groups. Although there are some collocates of *if you* shared by these three groups, the LIVT participants had more various collocates than the TESOL and NTESOL participants. From the patterns, it is found that the LIVT participants frequently used the word *like* in L1 position to introduce an illustration. We also found this kind of use of *like* in the COCAsp and BASEah corpora. However, the word *like* was not used in this way by the participants in the TESOL and NTESOL groups. This finding supports our first hypothesis: those EFL learners who go to the UK for study will show noticeable differences in their English use in comparison with those who stay in Taiwan. This finding suggests that the LIVT participants were likely influenced by the English input they encountered. It is also potentially compatible with what has been discussed in chapter 2. The LIVT participants had likely been primed differently from the TESOL and NTESOL participants in terms of the use of *like* with *if (you)*, and their primings for the use of *like* were possibly similar to the use found in the BASEah and COCAsp corpora, which are made up of texts from genres to which the LIVT participants had been exposed.

N	L1	Centre	R1
1	EH	IF YOU	WANT
2	YOU		DON'T
3	BUT		EH
4	SO		HAVE
5	BECAUSE		IF
6	LIKE		ARE
7	AND		LIKE
8	THINK		STUDY
9	IF		CAN
10	ENGLISH		GO
11	WHEN		YOU
12			JUST
13			USE

Patterns of *if you* in LIVT

N	L1	Centre	R1
1	BECAUSE	IF YOU	WANT
2	YOU		HAVE
3	OR		CAN
4	ENGLISH		LEARN
5			DEDICATE

Patterns of *if you* in TESOL

N	L1	Centre	R1
1	BUT	IF YOU	WANT
2	EH		ARE
3	IF		DON'T
4	THINK		KNOW
5			SAY
6			EH
7			GET
8			JUST

Patterns of *if you* in NTESOL

Table 4.36 Patterns of *if you* in LIVT, TESOL and NTESOL

N	Concordance
1	cific tips or ha(ve) &eh **like if you** read that in the readin
2	maybe you you interested **like if you** like baseball and you y
3	ike the view perspective **like if you** &eh when you &eh discu
4	ound silly &eh I thought **like if you** could speak English it

Table 4.37 Instances of *like if you* in LIVT

N	Concordance
1	looked like a chicken , **like if you** would take a bite out o
2	in most circumstances . **Like if you** 're interested in losin
3	has to be consequences . **Like if you** screw up my mother 's 4
4	t was that distracting . **Like if you** put on some ... @!GIFFO

Table 4.38 Instances of *like if you* in COCAsp

WHEN YOU

The LIVT participants used *when you* significantly more frequently than the TESOL and NTESOL participants, again supporting our first hypothesis. (See Table 4.39.)

when you	LLR	Sig.		
LIVT - TE	4.180	**0.041**	*	+
LIVT - NTE	7.228	**0.007**	**	+
TE - NTE	0.658	0.417		+
when	LLR	Sig.		
LIVT – TE	2.10	0.148		-
LIVT – NTE	0.37	0.541		+
TE – NTE	3.91	**0.048**	*	+

Table 4.39 Log-Likelihood Ratios: *when you* and *when*

The collocation *when you* is found once per 1000 words in the LIVT group, which is almost twice as frequent as in the TESOL group (0.54 /per 1000 words) and nearly three times more frequent than in the NTESOL group (0.39/per 1000 words). When examining the overall use of the word *when*, it is found that the most frequent R1 collocate of *when* is the pronoun *I*. There are 101 occurrences of *when I* found in the LIVT group (57% of the total use of *when*), 130 occurrences of *when I* in the TESOL group (61%), and 69 occurrences of *when I* in the NTESOL group (51%). The collocation *when you* occured 29 times in the LIVT group (16%), 16 times in the TESOL group (8%) and only 9 times in the NTESOL group (7%). The use of *when you* in the LIVT group is twice more frequent than the TESOL and NTESOL groups in proportion, and the reasons for this are again worth investigating. (See Table 4.40.)

	LIVT			TESOL			NTESOL		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WHEN	178	100	6.17	212	100	7.14	134	100	5.75
WHEN YOU	29	16	1.00	16	8	0.54	9	7	0.39
Other R1 collocates of *WHEN*	**I 101 (57%)** **YOU 29 (16%)** THEY 12 WE 8 WHEN 7 I'M 3 HE 3			**I 130 (61%)** WE 18 **YOU 16 (8%)** WHEN 12 THEY 6 EH 6 I'M 5			**I 69 (51%)** WE 13 WHEN 12 **YOU 9 (7%)** I'M 9 THEY 6 EH 4 THE 3		

Table 4.40 Frequency of use of *when* and *when you* and other collocates of *when* in LIVT, TESOL and NTESOL

When examining how the collocation *when you* was used by these participants, it is found that, similar to the use of *if you*, the pronoun *you* in this situation is often used as a general pronoun, which refers not to the interviewer, but to a general concept of people in order to state a general situation. This finding again reveals that the LIVT participants were prone to use the pronoun *you* differently from the TESOL and NTESOL participants.

N	Concordance
1	ty in listening but after **when you** are getting familiar with
2	and you have to speak and **when you** listen listen more the mo
3	een yeah I think would be **when you** go out you first talk wit
4	useful (...) yeah because **when you** can speak English in Taiw
5	words or grammar but but **when you** just grow up &eh , maybe

Table 4.41 Instances of *when you* in LIVT

When looking further at the collocation *when you* in these three groups, it is found that the LIVT participants tended to use the connective *but* in front of *when you,* and this kind of use is rarely found in the other two groups. (See Table 4.42.)

N	L1	Centre	R1
1	BUT	WHEN YOU	GO
2	YOU		ARE
3	YEAH		WHEN
4	LIKE		JUST
5			CAN
6			GOT

Patterns of *when you* in LIVT

N	L1	Centre	R1
1	THAT	WHEN YOU	ARE
2			LEARN
3			HAVE

Patterns of *when you* in TESOL

N	L1	Centre	R1
1	AND	WHEN YOU	
2			

Patterns of *when you* in NTESOL

Table 4.42 Patterns of *when you* in LIVT, TESOL and NTESOL

There are three instances of *but when you* found in the LIVT group (LIVT-1 to -3), and only one instance of *but when you* found in the TESOL group (TE-1) and none in the NTESOL group. The LIVT participants also tended to use discourse marker(s) *yeah (yes)* before *when you*, and such a use is not found in the TESOL and NTESOL groups. The use of the word *like* before *when you*, similar to the finding of *like if you* discussed previously, is also unique to the LIVT group.

LIVT-1 words or grammar but **but when you** just grow up &eh , maybe

LIVT-2 sometimes get confuse **but when you** see not only a a period b

LIVT-3 are about the grammar **but when you** contact with each other p

TE-1 ching second language **but when you** just use Chinese as your

LIVT-4 ix or six point five **yeah when you** are when they look your w

LIVT-5 rmal global language **yeah when you** go when you go any countr

LIVT-6 y difficult &haha . # **yes when you** met people you will find

Table 4.43 Instances of * *when you*

The TESOL participants also showed their own preference in using *when you*. The TESOL participant tended to use the connective *that* in front of *when you* (TE -2 to -3), and this kind of use is not found in the other two groups. Previously, when we investigated the collocations of *think* in the TESOL group, we also found that the TESOL participants were prone to use *that* as a connective in the collocation *think that*. This provides support for hypothesis 2. The finding of the preference to use *that* before *when you* in the TESOL group is potentially compatible with what has been discussed in chapter 2. It is possible that the TESOL participants might be primed to use the word *that* more

frequently as a connective and their primings were quite different from the LIVT and NTESOL participants' primings in this respect.

TE -2 se too so &eh I know **that when you** learn (..) a langu(age) a

TE -3 hat's very difficult **that when you** sound sounds more and rea

Table 4.44 Instances of *that when you* in TE

The NTESOL participants, although there are only 9 instances of *when you*, were also distinctive in their use of the collocation *when you*. The NTESOL participants tended to have the connective *and* in front of *when you*, and this collocation *and when you* is not found in the other two groups. There are 5 out of 9 instances of *when you* companied by the connective *and*. There is only one instance of *and when you* found in the LIVT group (one out of 29 instances of *when you*). There is only one instance of *and when you* in the TESOL group (one out of 16 instances of *when you*).

NTE -1 and when you &eh use **and when you** play more games and read

NTE -2 omposition of English **and when you** read reading or &eh in pu

NTE -3 t know the words yeah **and when you** &eh use and when you play

NTE -4 t know the words yeah **and when you** writing you you can't u(s

NTE -5 friends or classmate **and when you** speak more and you can an

Table 4.45 Instances of *and when you* in NTE

These differences suggest that when using the conditional *when you* as a means of expressing a general situation, the participants in the three groups behaved very differently. The LIVT participants tended to use *when you* to give a contrary scenario by using *but when you*, or to provide a relevant scenario in agreement with the previous speaker by using *yeah/yes when you*, while the TESOL participants and NTESOL participants seemed not to have such preferences. The TESOL participants used *that when you* to complement their sentences, and this kind of use is not found in the LIVT and NTESOL groups. The NTESOL participants used *and when you* more strongly than the LIVT and TESOL participants to add a further illustration. The collocations *but when you*, *and when*

you and *that when you* are all very common in the reference corpora. (See Table 4.46.) However, the distribution of use of *but when you*, *and when you* and *that when you* is very different from group to group. The use of *but when you*, *and when you* and *that when you* is evenly distributed in the two parallel corpora whereas, in the three groups, the distribution is very uneven. (See Table 4.47.) What we has found is that the three groups of participants had their own discourse preferences for answering the same questions. Although why the different preferences exists cannot be answered in the current study, this finding reveals a topic which may be worth studying in the future.

N	L1	Centre	R1
1	BUT	WHEN YOU	HAVE
2	SO		THINK
3	THAT		LOOK
4	YOU		START
5	IS		READ
6	KNOW		GO
7	AND		ACTUALLY

Patterns of *when you* in BASEah

N	L1	Centre	R1
1	AND	WHEN YOU	RE
2	BUT		LOOK
3	THAT		HAVE
4	KNOW		WERE
5	SO		GET
6	YOU		SEE

Patterns of *when you* in COCAsp

Table 4.46 Patterns of *when you* in BASEah and COCAsp

	LIVT		TE		NTE		BASEah		COCAsp	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
WHEN	178	6.17	212	7.14	134	5.75	881	2.08	46142	2.50
WHEN YOU	29	1.00	16	0.54	9	0.39	118	0.28	8539	0.46
BUT WHEN YOU	3	**0.10**	1	0.03	0	0	8	**0.02**	372	**0.02**
THAT WHEN YOU	0	0	2	**0.07**	0	0	7	**0.02**	347	**0.02**
AND WHEN YOU	0	0	0	0	5	**0.21**	3	**0.01**	564	**0.03**

Table 4.47 Use of the collocations of *when you* in all groups

WHAT YOU

The LIVT participants used *what* with the pronoun *you* much more often than the TESOL and NTESOL participants. However, it is found that the overall use of *what* does not differ greatly

between groups, therefore, the overuse of *what you* in the LIVT group seems to be particularly intriguing. (See Table 4.48.)

what you	LLR	Sig.		
LIVT – TE	15.092	**0.000**	***	+
LIVT – NTE	8.663	**0.003**	**	+
TE – NTE	0.63	0.429		-
what	LLR	Sig.		
LIVT – TE	3.98	**0.046**	*	+
LIVT - NTE	0.81	0.369		+
TE – NTE	0.96	0.327		-

Table 4.48 Log-Likelihood Ratios: *what you* and *what*

Examining the instances of the use of *what* and *what you*, it is found that this collocation fits it with our previous findings that the LIVT participants were potentially primed to use *you* as a general pronoun in their English more strongly than the TESOL and NTESOL participants. The amount of occurrences of the word *what* does not differ much across the three groups. However, the most common R1 collocates of *what* in these three groups, as the Table 4.49 shows, are somewhat different.

	LIVT			TE			NTE		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WHAT	109	100	3.78	84	100	2.83	77	100	3.30
WHAT YOU	15	**14**	**0.52**	1	**1**	**0.03**	2	**3**	**0.09**
Other R1 collocates of *WHAT*	THEY 29 **YOU 15** I 12 WHAT 6 THE 6 ARE 6 KIND 5			THEY 16 I 14 THE 11 WHAT 8 ARE 4 KIND 3 HE 3 EH 3			THEY 15 WHAT 11 THE 8 IS 5 I 5 I'M 4 EH 4 DO 4 HE 3		

Table 4.49 Frequency of use of *what you* and other collocates of *what* in LIVT, TESOL and NTESOL

There is only one instance of *what you* (TE-4) in the TESOL group and two instances in the NTESOL group (NTE-6 & -7), which does not permit any further analysis. On the other hand, there

are sufficient instances in the LIVT group to note that the LIVT participants used *understand what you* twice (LIVT-12 & -13), *what you say* twice (LIVT-11 & -12) and *what you want* four times. (LIVT -7 to -10). The NTESOL participants chimed with the LIVT participants' use of *understand what you* (NTE -7) and shared one instance with the LIVT participants by using the same expression *understand what you are talking about*.

N	L1	Centre	R1
1	UNDERSTAND	WHAT YOU	WANT
2			SAY

Table 4.50 Patterns of *what you* in LIVT

TE -4	ting is the production of **what you** think so if you write you
NTE -6	asy but not hard just &eh **what you** feel if &eh English is us
NTE -7	ey will try to understand **what you** are talking about and but
LIVT -7	y to try to let them know **what you want** to what do you want
LIVT -8	tell your tutors that is **what you want** to research or you w
LIVT -9	you just told your friend **what you want** to say and what you
LIVT-10	cal thinking and find out **what you want** and let you really g
LIVT -11	me they cannot just catch **what you say** for the idea they jus
LIVT -12	or your friend **understand what you say** and that is enough so
LIVT -13	ough they will **understand what you** are talking about you don

Table 4.51 Instances of * *what you*

In the above instances, the LIVT participants had more frequent use of *what you*, and had relatively more framed word sequences of *what you* than the TESOL and NTESOL groups. The LIVT participants used the pattern [*verbs*] *what you* [*verbs*] in their spoken English for referencing the idea/object more strongly than the TESOL and NTESOL participants. This may result from the influence of their authentic English input, because in the BASEah and COCAsp corpora, the collocations *what you want* and *what you said* appear to be very common uses in English. (See Table

4.52.) It is possible that the LIVT participants had heard of this kind of usage frequently in their daily lives, and therefore, had been influenced to use the pattern of *what you* more strongly than the TESOL and NTESOL particpants. What we had found supports for our hypothesis 1, that the EFL learners who go to the UK for study will show noticeable differences from the ones in Taiwan. This finding also implies the different tendency of applying this usage more strongly by the LIVT participants might result from the English input they had received. The findings are likely compatible with what has been discussed in chapter 2. When the use of words occurs frequently, it is likely to become stored in a person's long-term memory and easier to be retrieved afterwards. Therefore, people are likely to be strongly primed to use the words/combination of words to which they are frequently exposed. What we have observed suggests that the immersion in different language environments is very likely to cause people to be primed differently in their language use.

N	L1	Centre	R1
1	AND	**WHAT YOU**	KNOW
2	OF		HAVE
3	IS		**WANT**
4	TO		DO
5	KNOW		SEE
6	SO		MIGHT
7	ON		COULD
8	DO		CHOOSE
9	ABOUT		THINK
10	THAT		WERE

Patterns of *what you* in BASEah

N	L1	Centre	R1
1	KNOW	WHAT YOU	RE
2	IS	WHAT ? YOU	DO
3	OF	WHAT . YOU	THINK
4	AND		HAVE
5	THAT		**WANT**
6	ABOUT		**SAID**
7	DO		VE
8	TO		CAN
9	THINK		WERE

Patterns of *what you* in COCAsp

Table 4.52 Patterns of *what you* in BASEah and COCAsp

4.3.2 R1 Collocates of *you*

In the investigation of the R1 collocates of *you*, we found that the modal verbs *can* and *will* all appear in the top 10 most frequent R1 collocate lists of the three groups. The verbs *want*, *mean*, *have* and *know* also appear in all three groups' top 10 most frequent R1 collocate lists. The auxiliary verb *are* and the adverb *just* only appear in the LIVT and NTESOL groups' lists. The word *to* only appears in the LIVT group's list. The auxiliary verbs *do('nt)* and the verbs *think* and *learn* only appear in the TESOL group's list. (See Table 4.44.)

	LIVT			TESOL			NTESOL		
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
--	--	714	24.73	--	299	10.07	--	236	10.13
1	CAN	78	2.70	HAVE	42	1.41	CAN	32	1.37
2	WILL	49	1.70	CAN	41	1.38	HAVE	21	0.90
3	JUST	44	1.52	KNOW	29	0.98	MEAN	20	0.86
4	YOU	39	1.35	DON'T	17	0.57	KNOW	20	0.86
5	KNOW	38	1.32	MEAN	14	0.47	WANT	13	0.56
6	HAVE	37	1.28	WANT	10	0.34	JUST	12	0.51
7	WANT	35	1.21	YOU	8	0.27	YOU	9	0.39
8	TO	30	1.04	LEARN	8	0.27	ARE	9	0.39
9	MEAN	26	0.90	WILL	6	0.20	WILL	7	0.30
10	ARE	22	0.76	THINK	6	0.20	EH	6	0.26

Table 4.53 R1 Collocates of *you* in LIVT, TESOL and NTESOL

Based on the log-likelihood ratios of these collocations (see Table 4.54), it is found that the LIVT participants used the collocations *you will*, *you to*, *you want* and *you can* significantly more frequently than the TESOL and NTESOL participants. The LIVT and NTESOL participants used the collocation *you just* significantly more frequently than the TESOL participants. As we had discussed previously, the significantly more frequent use of the collocations in the LIVT group is very likely due to that fact that the LIVT participants overly applied *you* as a general pronoun for illustration and generalisation. The highly frequent occurrences of these collocations imply the LIVT participants were strongly primed to use these collocations in their spoken English. The overly use of *you just* and *you will* in the LIVT group, as we had discussed in the use of *I just* and *I will*, is possibly resulted from the fact that the LIVT participants tended to apply the word *just* more often for expressing hedging and *will* for various discourse purposes. The collocations *you can* and *you want* are used to talk about what a person intends to do and what s/he can achieve. The overly use of these two collocations in the LIVT group is likely relevant to the different focuses that the LIVT participants

had in their opinions from the TESOL and NTESOL participants. Although it would be interesting to see how these participants used *you just*, *you can* and *you want*, it is impossible to cover all the differences between groups in this study due to the space of this thesis. Therefore, we selected the collocation *you to* which seemingly have noticeable differences for our test of hypotheses.

LIVT - NTE					LIVT - TE					TE - NTE				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.							
YOU WILL	27.081	0.000	***	+	*YOU JUST*	40.129	0.000	***	+					
YOU TO	**20.238**	**0.000**	*******	**+**	*YOU WILL*	39.547	0.000	***	+					
YOU YOU	14.340	0.000	***	+	YOU YOU	23.143	0.000	***	+					
YOU JUST	13.228	0.000	***	+	**YOU TO**	**20.515**	**0.000**	*******	**+**					
YOU CAN	11.249	0.001	***	+	YOU LEARN	20.061	0.000	***	+					
YOU WANT	6.308	0.012	*	+	YOU WANT	15.414	0.000	***	+					
					YOU ARE	14.225	0.000	***	+					
					YOU CAN	12.750	0.000	***	+					
					YOU MEAN	3.998	0.046	*	+					
Less					Less					Less				
										Words	LLR	Sig.		
										YOU JUST	6.355	0.012	*	-

Table 4.54 Log-Likelihood Ratios: The R1 collocates of *you* in LIVT, TESOL and NTESOL

YOU TO

The collocation *you to* was significantly more frequently used by the LIVT particpants than by the TESOL and NTESOL participants. There are 30 occurrences of *you to* found in the LIVT group, while there are only 5 occurrences found in the TESOL group and 3 occurrences in the NTESOL group. (See Table 4.55.) In the 30 instances of the LIVT group, the collocation *you to* was used in the patterns of (a) [*for you to + verb*] and (b) [*verb + you to + verb*]. It is found that the LIVT participants used both patterns in their speech, but such patterns were rarely found to be used by the TESOL and NTESOL participants.

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
YOU TO	30	1.04	5	0.17	3	0.13
Patterns of *you to*	N L1 Centre R1 1 FOR YOU TO LEARN 2 ASK EH 3 HELP UNDERSTAND 4 FORCED USE 5 STUDY 6 PASS 7 SPEAK		N L1 Centre R1 1 FOR YOU TO 2			
L1 Collocates of *you to*	FOR 8 ASK 6 HELP 3 FORCED 2 WHAT 1 TRAIN 1 FUN 1		FOR 3 ENCOURAGE 1 PUSH 1		ASK 1 TEACH 1 WANT 1	

Table 4.55 Use and Patterns of *you to* in LIVT, TESOL and NTESOL

When taking a closer look at the patterns, it is found that there are two kinds of patterns involved in the use of *you to*. The first kind is the pattern [*for you to + verb*]. The other kind is the pattern [*verb + you to + verb*]. Here we will investigate the first kind. There are 8 instances of *for you to* found in the LIVT group (LIVT-14 to -20), 3 instances found in the TESOL group, but none in the NTESOL group. There are two kinds of L1 colligates (grammatical units) appearing in front of this pattern [*for you to + verbs*]. One is the adjectives, such as *easy*, *strange* or *good*. Four out of 8 instances of *for you to* in the LIVT group are accompanied with an adjective in the L1 position (LIVT-14 to -17). There are two instances of [*adjective + for you to*] in the TESOL group, and these two instances are in the form of *it's easy for you to* (TE -5 & -6). The other kind of L1 colligates of *for you to* is nouns. Three out of 8 instances of *for you to* in the LIVT group are in the pattern [*noun + for you to*] (LIVT-18 to -20). One out of three instances of *for you to* in the TESOL group is accompanied by a noun (TE-7).

LIVT-14	y in the U_K it's easy **for you to** travel around &haha to Eur
LIVT-15	d but it's &eh strange **for you to** understand yeah . # yeah b
LIVT-16	you it's quite helpful **for you to** learn English . # &eh when
LIVT-17	t (..) that's not good **for you to** use in the real life , yes
LIVT-18	e many &eh information **for you to** explore for you to search
LIVT-19	nce , an opportunities **for you to** use language or you will b
LIVT-20	, there are two years **for you to** study this program , but i
LIVT-21	ion for you to explore **for you to** search , yes . # yes . # w
TE-5	eful because it's easy **for you to** forget this (be)cause mayb
TE-6	ersation and it's easy **for you to** make friends yeah . # oh I
TE-7	but it's a it's a push **for you to** polish your skills . # res

Table 4.56 Instances of * *for you to*

Here we can see how different the participants of the three groups are. The LIVT participants tended to use *for you to* more often than the TESOL and NTESOL participants, and the way they used *for you to* in their speech is more flexible. Unlike the relatively more fixed use of *easy for you to* found in the TESOL group, the LIVT group shows various adjectives applied in the pattern [*adjective + for you to*]. A similar situation occurs in the use of the pattern [*noun + for you to*], the LIVT participants showed more flexibility in the use of *for you to*. The TESOL participants showed less flexibility of the use, and the NTESOL participants simply did not use this colligation in their spoken English at all. In chapter 2, we has discussed that non-native ESL speakers merely recognise/use word strings to which they have been frequently exposed and native speakers do not merely recognise/use them. They tend to be more open to process flexible forms of word strings (Ellis et al., 2009). The LIVT participants are not native speakers. However, in comparison with the TESOL and NTESOL participants, the LIVT participants had received more authentic English input from native speakers. If as discussed in chapter 2, input is the key element to influence people's language use, this kind of differences between the LIVT participants and the TESOL/NTESOL participants is the expected result. Moreover, the openness to use more flexible patterns/formulaic sequences in their language use indicates that the LIVT participants' priming of language becomes similar to the priming of native

speakers. This finding to some degree echoes to what Hoey (2005), Ellis et al. (2009) and Pace-Sigge (2013) suggested in their studies.

The second kind of use of *you to* is the pattern [*verb + you to*]. There are 12 instances of [*verb + you to*] found in the LIVT group, and these verbs comprise *ask* (6 occurrences), *help* (3 occurrences), *forced* (2 occurrences) and *train* (1 occurrence). There are only 2 instances of [*verb + you to*] in the TESOL group and the verbs are *encourage* (1 occurrence) and *push* (1 occurrence) (TE-8&9). There are 3 instances in the NTESOL group and the verbs are *ask* (1 occurrence), *teach* (1 occurrence) and *want* (1 occurrence). (See Tables 4.57 and 4.58.)

TE-8 m environment to encourage **you to** use the language so &eh so

TE-9 # &eh (...) they will push **you to** study English just for hig

NTE-8 although teacher will ask **you to** oh maybe you can watch &eh

NTE-9 se &eh public school teach **you to** to read the books but cram

NTE-10 h they they want they want **you to** get maybe G_E_P_T or &eh &

Table 4.57 Instances of * *you to* in TE and NTE

	LIVT			TE			NTE		
Patterns	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
you to	30	--	1.04	5	--	0.17	3	--	0.13
for you to + verb	8	100	**0.28**	3	100	**0.10**	0	0	**0.00**
adj. + for you to + verb	4	50	0.14	2	67	0.07	0	0	0.00
n.+ for you to + verb	3	38	0.10	1	33	0.03	0	0	0.00
verb + you to + verb	12	--	**0.42**	2	--	**0.07**	3	--	**0.13**

Table 4.58 Use of Patterns of *you to* in LIVT, TESOL and NTESOL

We can conclude that the LIVT participants tended to use *you to* more strongly and their use of *you to* was very different from the TESOL and NTESOL participants. The findings support for our hypothesis 1 that the EFL learners studying in the UK will have noticeable differences from the EFL

learners studying in Taiwan. When checking the BASEah and COCAsp reference corpora, it is found that there are a wide range of collocates/colligates, as the table below shows.

N	L2	L1	Centre	R1
1	TO	WANT	YOU TO	DO
2	I'D	FOR		THINK
3	WE	ASKING		READ
4	WE'RE	ASK		GO
5	I'M	LIKE		THE
6	DON'T	GET		BE
7	YOU	ENCOURAGING		MAKE
8	WOULD	HELP		TO
9	WILL	INTRODUCE		FOCUS
10	JUST	ENCOURAGE		THAT
11	I'LL	RECOMMEND		COMPOSE
12	THAT	TO		WRITE
13	PARTICULARLY	ALERT		CHOOSE
14	EXPECT	MEMORISE		
15	LEAVE	LOOK		
16	ASKED	NOTICE		

Patterns of *you to* in BASEah

N	L2	L1	Centre	R1
1	TO	WANT	YOU TO	BE
2	WE	FOR		DO
3	N'T	ASK		THE
4	THEY	NEED		KNOW
5	JUST	OF		GET
6	IT	THANK		GO
7	THAT	LIKE		TAKE
8	WOULD	TAKE		COME
9	WILL	ASKED		SAY
10	AND	GET		HAVE
11	HE	WANTS		TELL
12	WHAT	TO		SEE
13	TIME	ASKING		MAKE
14	LL	ALLOW		THINK
15	LIKE	WANTED		LISTEN

Patterns of *you to* in COCAsp

Table 4.59 Patterns of *you to* in BASEah and COCAsp

The differences between the LIVT participants and the participants of the other groups, especially the NTESOL participants, suggest that the English input the LIVT group had received might play a crucial role in their language acquisition. It is possible that the LIVT participants had heard (input) and used (output) these patterns more frequently than the other precipitants in their environment. Therefore, their English use was influenced accordingly. That is, their priming of language is likely to be affected by their input as has been discussed in chapter 2.

4.3.3 Summary of the use of *you*

In 4.3, we investigated the top 10 most frequent collocates of the pronoun *you* in the LIVT, TESOL and NTESOL groups. We investigated several collocates of *you* that showed noticeable differences in the frequencies of the use between the three groups. What we found support our first and second hypotheses that the EFL learners with different language learning backgrounds will behave differently in their language use. It also suggests that different language learning backgrounds may influence not only the grammatical and lexical aspects of learners' English use, but very possibly the discourse aspects of their language use as well. The findings of the investigation of the use of *you* are presented in Table 4.60.

Support for Hypothesis 1		
Differences	Possible Causes of the differences	Words
The LIVT participants used the words **significantly more frequently** than the TESOL and NTESOL participants.	Input Influence	*what you; (like)if you; you to*
	Different discourse preferences	*when you*
Support for Hypothesis 2		
Differences	Possible Causes of the differences	Words
The use of *that* as connective with *when you* in the TESOL group is **rarely found in the other groups**.	Input Influence	*when you*
The TESOL participants used the words **significantly more frequently** than the NTESOL participants.	Different discourse preferences	*if you*

Table 4.60 Findings in the investigation of the use of *you*

Several differences are found to support our first and the second hypotheses. With regard to the use of the collocation *if you*, the LIVT participants tended to use *if you* significantly more often than the TESOL and NTESOL participants, and they used the word *like* in L1 position to introduce an illustration more frequently than the TESOL and NTESOL participants. This finding is potentially compatible with what has been discussed in chapter 2. The use of *like* with *if (you)* is similar to the use in the BASEah and COCAsp corpora, which are made up of texts from genres to which the LIVT participants had been exposed. When the use of words occurs frequently, it is likely to become stored in a person's long-term memory and easier to be retrieved afterwards. Therefore, people are likely to be strongly primed to use the words/combination of words to which they are frequently exposed. We also found that the TESOL participants tended to use the connective *that* in front of *when you*, and this kind of use is not found in the other two groups. When we investigated the collocations of *think* in the TESOL group, we found that the TESOL participants were primed to use *that* as a connective in the collocation *think that* more strongly than the other participants. The finding of this preference to use *that* before *when you* in the TESOL group suggests that the TESOL participants may have different priming of the use of *that*. This provides support for our hypothesis 2.

The significantly more frequent use of *when you* in the LIVT group is likely to be a result of different discourse behaviours. The LIVT participants used the conditional *when you* as a means of expressing a general situation more often than the participants in the other groups. The LIVT participants used *what* with the pronoun *you* much more often than the TESOL and NTESOL participants, and they used *what you* in a relatively fixed pattern. The strong tendency to use the pattern *verb + what you + verb* in the LIVT group might result from the influence of their authentic English input. (This kind of patterns is commonly found in the BASEah and COCAsp corpora.) A similar situation can also be found with the use of *you to*. The collocation *you to* was significantly more frequently used by the LIVT participants than by the TESOL and NTESOL participants. In the case of *you to*, the LIVT participants were found to be primed to use a wider variety of collocates and colligates, while the TESOL and NTESOL participants were not yet primed this way. All of these findings show that the differences in the langauge use among learners with different immersion/educational backgrounds exist, and different input is likely to be the cause.

4.4 The Use of Pronoun *it*

At the very beginning of this chapter, we found that the impersonal pronouns *it* and 它 [it] were used differently by these groups. The NTESOL participants used the impersonal pronoun 它 [it] in Mandarin more often than the participants of the other two groups, and the LIVT and TESOL participants used the English impersonal pronoun *it* more than the NTESOL participants. However, when checking the log-likelihood ratios of the uses of *it* and 它 [it], it is found that the frequency of the English impersonal pronoun *it* does not show great differences. No group shows a strong tendency of using *it* in their spoken English overly. In the case of the Mandarin pronoun 它 [it], the NTESOL participants, however, did reveal a tendency of using the word 它 [it] significantly more frequently in their spoken Mandarin. (See Table 4.61.) In the following section, we will investigate the collocations of *it*. The use of Mandarin 它 [it] will be discussed in the next chapter.

		Freq.	Per 1000 words		LLR	Sig.		
IT	LIVT	267	9.25	LIVT-TE	1.330	0.249		+
	TESOL	248	8.35	LIVT-NTE	0.485	0.486		+
	NTESOL	202	8.67	TE-NTE	0.151	0.698		-
		Freq.	Per 1000 words		LLR	Sig.		
它 (tā)	LIVT	331	8.69	LIVT-TE	0.067	0.796		+
	TESOL	296	8.51	**LIVT-NTE**	26.400	**0.000**	***	-
	NTESOL	515	12.43	**TE-NTE**	27.762	**0.000**	***	-

Table 4.61 Log-Likelihood Ratios: *it* and 它 (tā) in LIVT, TESOL and NTESOL

4.4.1 *L1 Collocates of it*

In the lists of the top 10 most frequent L1 collocates of English pronoun *it*, similar to the pronouns *I* and *you*, the filler sound *eh* and the repetition of the pronoun *it* are shared across all three groups. The connectives *because*, *but*, and *so* are also found in all three groups' lists. The connective *and* is only found in the LIVT and TESOL groups' top 10 most frequent L1 collocate lists. Many verbs are in the L1 position of *it*. (See Table 4.62.) When examining whether the use of these collocations differs significantly across the three groups, it is found that there are few L1 collocates showing significant differences in their frequencies. (See Table 4.63.)

	LIVT	267	9.25	TESOL	248	8.35	NTESOL	202	8.67
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	THINK	40	1.39	EH	27	0.91	THINK	18	0.77
2	LIKE	22	0.76	THINK	16	0.54	LEARN	15	0.64
3	IT	19	0.66	LIKE	16	0.54	IT	14	0.60
4	EH	14	0.48	IT	10	0.34	EH	13	0.56
5	SO	13	0.45	AND	10	0.34	USE	10	0.43
6	YEAH	7	0.24	SO	8	0.27	LIKE	9	0.39
7	TO	7	0.24	BECAUSE	8	0.27	BUT	8	0.34
8	BUT	7	0.24	THAT'S	7	0.24	BECAUSE	8	0.34
9	BECAUSE	7	0.24	SAY	6	0.20	SO	5	0.21
10	AND	7	0.24	BUT	6	0.20	READ	5	0.21

Table 4.62 L1 collocates of *it* in LIVT, TESOL and NTESOL

The LIVT participants used the collocations *think it*, *yeah it* and *to it* significantly more frequently than the TESOL and NTESOL participants. We will investigate how the participants used the collocations *to it* and *think it* to test our hypothesis 1. In the previous sections, we investigated the use of *yeah I* and found that the LIVT participants tended to use *yeah* more often in their spoken English. Therefore, we will not investigate the use of *yeah it* further in this section. In Table 4.52, we also found that the TESOL participants used the collocations *use it* and *learn it* significantly less frequently than the NTESOL participants. The LIVT participants also used the collocation *learn it* significantly less frequently than the NTESOL participants. The use of collocation *use it* will be investigated for our test of hypothesis 2. The use of *learn it* will be investigated for both hypotheses1 and 2.

LIVT - NTE					LIVT - TE					TE- NTE				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
THINK IT	4.510	**0.034**	*	+	**THINK IT**	11.305	**0.001**	***	+					
YEAH IT	3.869	0.049	*	+										
TO IT	3.869	0.049	*	+										
Less					Less					Less				
Words	LLR	Sig.								Words	LLR	Sig.		
LEARN IT	14.230	**0.000**	***	-						**LEARN IT**	14.646	**0.000**	***	-
										USE IT	10.886	0.001	***	-

Table 4.63 Log-Likelihood Ratios: The L1 collocates of *it* in LIVT, TESOL and NTESOL

TO IT

The collocation *to it* was not used very often by the three groups. There are 7 occurrences of *to it* in the LIVT group, 5 occurrences of *to it* in the TESOL group and only 1 occurrence of *to it* in the NTESOL group. When checking the instances in these three groups, it is found that the collocation *to it* is often used in the pattern [*get/got/BE-verb + used to it*]. There are 4 occurrences of the use of this pattern [*get/got/BE-verb + used to it*] in the LIVT group, 4 occurrences of this use in the TESOL group and 1 occurrence of this use in the NTESOL group. (See Table 4.64.) In the other use of *to it*, the word *to* is used as a preposition or an infinitive-*to*, and the speakers pause and start a new sentence with *it*. (See the instances of these two usages below Table 4.66.) These findings support our hypothesis 1 that the EFL learners who go to the UK for study will have noticeable differences from the ones in Taiwan. Although all the participants in the three groups tended to use the collocation *to it* in a fixed pattern [*get/BE-verb used to it*], the LIVT participants used the collocation *to it* more often when talking about how their habits were formed than the NTESOL participants. This finding shows that the LIVT participants behaved differently in discourse behaviours from the participants in the other two groups.

Patterns	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
to it	7	0.24	5	0.17	1	0.04
get/BE-verb used to it	**5**	**0.17**	4	0.13	**1**	**0.04**
to(prep/infinitive) <Pause> *it* (*subject)*	2	0.07	1	0.03		

Table 4.64 Use of *to it* in LIVT, TESOL and NTESOL

N	Concordance
1	yes . # &eh (...) it's easy **to it** helps u(s) the the &ch tea
2	it's okay but you were used **to it** so every time you just wat

Table 4.65 Instances of *to it* in LIVT

THINK IT

In our investigation on the use of collocation *think it*, we found that the LIVT participants used the collocation more frequently than the TESOL and NTESOL participants. (See Table 4.66.) There are 40 occurrences of the uses of *think it* (1.39 occurrences per 1000 words) in the LIVT group. There are 16 occurrences of *think it* (0.54 occurrences per 1000 words) in the TESOL group and the 18 occurrences of *think it* (0.77 occurrences per 1000 words) in the NTESOL group. The LIVT participants used the collocation *think it* almost twice more often than the NTESOL participants and nearly three times more often than the TESOL participants. The collocation *think it* is mainly accompanied with the pronoun *I*. 85% of the use of *think it* in the LIVT group is with *I*. In the TESOL and NTESOL groups, the use of *I think it* accounts for nearly 100% of the instances of *think it*. The highly frequent use of *think it* in the LIVT group and the strong link of *think it* with *I* echo our previous observation on the use of *I think*. We found that the LIVT participants were prone strongly to use the collocation *I think* as a starter to their sentences, while the TESOL and NTESOL participants, although they also used this collocation, were not prone as strongly as the LIVT participants were. (See the analysis of *I think* in 4.2.2.)

	LIVT			TE			NTE		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
THINK	599	100	20.75	478	100	16.10	406	100	17.42
THINK IT	40 (100%)	7	**1.39**	16 (100%)	3	**0.54**	18 (100%)	4	**0.77**
L1 Collocates of *THINK IT*	I 34(85%) **DON'T 3(5%)** WILL 1 HAVE 1			I 15 (94%) YOU 1			I 18(100%)		

Table 4.66 Use of *think* and *think it* in LIVT, TESOL and NTESOL

When taking a closer look at the patterns of use of *think it*, it is found that the LIVT participants used the collocation *think it* in both affirmative and negative sentences. (See Table 4.67.) There are 3 instances of *I don't think it* (5% of the total uses of *think it*) found in the LIVT group and there is none found in either of the TESOL and NTESOL groups. The findings support our hypothesis 1 that the EFL learners who go to the UK for study will have noticeable differences from those in Taiwan. The LIVT participants tended to use *I think/I don't think* as a strategy for starting a sentence. The TESOL and NTESOL participants did not use this discourse strategy as strongly as the LIVT participants. If the TESOL and NTESOL participants did use, they did not use it in the negative form.

N	L1	Centre	R1
1	I	THINK IT	WILL
2	DON'T		IS
3			WAS
4			DEPENDS
5			IT
6			WOULD
7			IT'S
8			SHOULD

Patterns of *think it* in LIVT

N	L1	Centre	R1
1	I	THINK IT	IS
2			DEPENDS
3			

Patterns of *think it* in TESOL

N	L1	Centre	R1
1	I	THINK IT	IS
2			IT

Patterns of *think it* in NTESOL

Table 4.67 Patterns of *think it* in LIVT, TESOL and NTESOL

LEARN IT

In our investigation on the use of collocation *learn it*, it is found that the NTESOL participants had a tendency to use this collocation more often than the LIVT and TESOL participants . (See Table 4.68.) There are 15 occurrences of the use of *learn it* (0.64 occurrences per 1000 words) in the NTESOL group. There are mere 3 occurrences of *learn it* (0.1 occurrences per 1000 words) in the LIVT group and only 2 occurrences of *learn it*. (0.07 occurrences per 1000 words) in the TESOL group. The NTESOL participants used the collocation *learn it* 60 times more often than the participants in the other two groups. When we have a further look at the use of *learn it*, another interesting tendency is found. Many instances of *learn it* in the NTESOL group reveal an association with reluctant obligation. There are 7 instances of *to learn it* in the NTESOL group, and 5 of them are in such forms as *have to learn it, have… pressure to learn it* and *be forced to learn it*. There was another collocation *should learn it* found in the NTESOL group which also has a sense of reluctance. (See Table 4.69.) Such a usage is not found so explicitly in the LIVT and TESOL groups. Although the finding supports our hypothesis that the EFL learners who go to the UK for study will behave differently in their language use from those in Taiwan, and hypothesis 2 that the EFL learners who study in English-relevant subjects will behave differently from the ones in non-English-relevant subjects, the cause of these noticeable differences seems not to derive from their differences in language input but from these participants' different attitudes to learning. Our finding suggests that the NTESOL participants tended to talk about learning English (it) more frequently than the LIVT and TESOL participants. Moreover, when using the collocation *learn it (English)*, the NTESOL participants were more likely to use it to express a sense of unwillingness. This finding does not suggest that the NTESOL participants were influenced by lack of the external input from the environment or learning materials. What this finding shows is that the NTESOL participants' anxiety and unease in English learning were greater than the LIVT and TESOL participants.

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
LEARN	154	100	5.33	136	100	4.58	115	100	4.93
LEARN IT	3	**1.95**	**0.10**	2	**1.47**	**0.07**	15	**13.04**	**0.64**

Table 4.68 Use of *learn* and *learn it* in LIVT, TESOL and NTESOL

N	Concordance
1	it's kind of be **forced to learn it** because if if you don't l
2	n school so so we **have to learn it** &eh but for me I think an
3	ust don't like we **have to learn it** all for the test and afte
4	many resource you **have to learn it** yourself or use the resou
5	) &eh it's interesting to learn it but I don't like the stre
6	a little bit **pressure to learn** it &eh . # &eh because &eh I
7	&eh spend much time to to learn it so &eh I think &eh if I g
8	in languages so we **should learn it** if you want to get more a

Table 4.69 Instances of *learn it* from NTESOL

USE IT

Previously, we found that the TESOL participants used the collocation *use it* significantly less frequently than the NTESOL participants. The TESOL participants used the collocation *use it* twice (0.07 times per 1000 words) and the NTESOL participants used the collocation *use it* 10 times (0.43 times per 1000 words.) The LIVT participants also used the collocation *use it* more frequently than the TESOL participants, but the differences were not as great as with the NTESOL participants. (See Table 4.70.)

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
USE	83	100	2.87	76	100	2.56	78	100	3.35
USE IT	6	7.23	**0.21**	2	2.63	**0.07**	10	12.82	**0.43**
HOW TO USE	7	8.43	0.24	7	9.21	0.24	1	1.28	0.04
Patterns of *use it*	N L2 L1 Centre R1 1 **HOW** **TO** **USE IT** IN						N L2 L1 Centre R1 1 **TO** **USE IT** YEAH 2 IN		
Patterns of *use*	N L2 L1 Centre R1 1 YOU **TO** **USE** THE 2 **HOW** WILL ENGLISH 3 THEY CAN IN 4 HAVE THEY **IT** 5 WILL YOU TO			N L2 L1 Centre R1 1 **HOW** **TO** **USE** THE 2 YOU CAN ENGLISH 3 LIKE THEY THAT 4 HAVE WILL THIS 5 USED DON'T EH			N L2 L1 Centre R1 1 YOU TO **USE** ENGLISH 2 HAVE CAN THE 3 WE EH **IT** 4 CAN CAN'T IN 5 EH WE EH		
Patterns of *how to use*	N L2 L1 Centre R1 1 DON'T KNOW **HOW TO USE** **IT** 2 **THE**			N L2 L1 Centre R1 1 DON'T KNOW **HOW TO USE** **THE**					

Table 4.70 Use of *use*, *use it* and *how to use* in LIVT, TESOL and NTESOL

When investigating the patterns of the word *use*, it is found that the TESOL participants tended not to use the pronoun *it* to substitute for the things they said. The word *it* was used very often with the verb *use* by the LIVT and NTESOL participants in the patterns presented in Table 4.70. However, the TESOL participants seldom used this usage. This tendency can be seen more clearly in the observation of the use of a common expression by the groups. In the use of *use* and *use it*, we can see that the pattern [*how to use (it)*] was commonly used by the LIVT and TESOL participants. (There is only one occurrence of *how to use (it)* in the NTESOL group.) The expression *how to use* is often in circumstances where both the participants and I knew what was being referred to in our conversations. (e.g. the learning materials, difficult vocabulary) However, when checking the use of *how to use* in these groups, it is found that the TESOL still tended not to use the pronoun *it* to substitute for an item in our talk. The TESOL group tended to use the item name when they used the verb *use*. (See Table 4.71.) This strong tendency was unique to the TESOL participants. Why the TESOL participants had this special preference in the use of the verb *use* and whether this preference was influenced by their English-major educational backgrounds are difficult to tell. But still, this finding supports for our hypothesis 2 that there are noticeable differences between the English-major EFL learners from those without similar backgrounds.

1	k it's difficult and and how to use **how to use the words** acc
2	word but you don't know **how to use this word** so maybe after
3	ord it is you don't know **how to use the precise word** to desc
4	but I don't exactly know **how to use the words** as for example
5	want to &eh really know **how to use the English** the function
6	icult and and how to use **how to use the** words accuracy is re
7	happy but you don't know **how to use that** so I think &eh now

Table 4.71 Instances of *how to use* in TESOL

4.4.2 R1 Collocates of *it*

In the lists of the top 10 most frequent R1 collocates of English pronoun *it*, like what we found in the pronouns *I* and *you*, the filler sound *eh* and the repetition of the pronoun *it* are common in all three groups. (See Table 4.72.) The modal verb *will* is in all the three groups' lists, but the frequency of *it will* varies greatly from group to group. There are many R1 collocates that only appear in one of the groups' top 10 most frequent R1 collocate lists. The verb *depends* and the modal verb *would* are in the LIVT group's top 10 most frequent list. The verb *has* is only in the TESOL group's top 10 most frequent list, and the connective *because* is only in the NTESOL group's top 10 most frequent list. Although it seems that these three groups have very different R1 collocates of *it*, when we examine whether the use of these collocations has statistical significance in terms of their occurrences across the three groups, the differences are found to be significant but quite weakly so in four cases. (See Table 4.73.) The use of the collocation *it was* in the NTESOL group is significantly less frequent in comparison with the one in the LIVT and TESOL groups. This finding justifies our previous finding that the NTESOL participants were not prone strongly to use past tense verbs/auxiliary verbs in their spoken English in 4.2.2. The use of the collocation *it will* in the LIVT group, however, shows an interesting tendency to be significantly heavily used in comparison with the TESOL and NTESOL groups. We go on to examine whether there are any noteworthy features of the collocation *it will* in the next section.

	LIVT	267	9.25	TESOL	248	8.35	NTESOL	202	8.67
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	**WILL**	**27**	**0.94**	IS	33	1.11	IS	27	1.16
2	IS	26	0.90	WAS	20	0.67	EH	22	0.94
3	IT	19	0.66	EH	16	0.54	IT	14	0.60
4	EH	19	0.66	**WILL**	**11**	**0.37**	SO	10	0.43
5	YEAH	11	0.38	SO	10	0.34	YEAH	8	0.34
6	IT'S	10	0.35	IT	10	0.34	CAN	6	0.26
7	WAS	9	0.31	AND	8	0.27	**BECAUSE**	6	0.26
8	DEPENDS	9	0.31	HAS	7	0.24	**WILL**	**5**	**0.21**
9	WOULD	6	0.21	IT'S	6	0.20	JUST	5	0.21
10	SO	6	0.21	YEAH	5	0.17	BUT	5	0.21

Table 4.72 R1 collocates of *it* in LIVT, TESOL and NTESOL

LIVT - NTE					**LIVT - TE**					**TE- NTE**				
Over					Over					Over				
Words	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
IT WILL	**12.279**	**0.000**	*******	**+**	**IT WILL**	**7.404**	**0.007**	******	**+**	IT WAS	16.781	0.000	***	+
IT WAS	5.763	0.016	*	+										
Less					Less					Less				
					Word	LLR	Sig.			Word	LLR	Sig.		
					IT WAS	3.978	0.046	*	-	IT JUST	3.967	0.046	*	-
										IT BUT	3.967	0.046	*	-

Table 4.73 Log-Likelihood Ratios: The R1 collocates of *it* in LIVT, TESOL and NTESOL

IT WILL

The use of *it will* and *will* is found to be significantly more frequent in the LIVT group than in the TESOL and NTESOL groups. (See Table 4.74.)

it will	LLR	Sig.		
LIVT - TE	7.404	**0.007**	******	**+**
LIVT - NTE	12.279	**0.000**	*******	**+**
TE - NTE	1.088	0.297		+
will	LLR	Sig.		
LIVT - TE	5.68	**0.017**	*****	**+**
LIVT - NTE	13.18	**0.000**	*******	**+**
TE - NTE	2.01	0.156		+

Table 4.74 Log-Likelihood Ratios: *it will* and *will* in LIVT, TESOL and NTESOL

There are 27 occurrences of *it will* in the LIVT group (0.94 occurrences per 1000 words), 11 occurrences of *it will* in the TESOL group (0.37 occurrences per 1000 words) and only 5 occurrences of *it will* in the NTESOL group (0.21 occurrences per 1000 words.) The frequency of *it will* in the LIVT group is almost 3 times more than in the TESOL and NTESOL groups. (See Table 4.75.) The LIVT participants also used the word *will* significantly more frequently than the TESOL and NTESOL participants. However, the frequency of the word *will* in the LIVT group and in the other two groups does not differ as greatly as the one of the use of *it will*. The LIVT participants used the word *will* 7.24 times per 1000 words. The TESOL participants used the word *will* 5.66 times per 1000 words and the NTESOL participants used the words *will* 4.76 times per 1000 words. The LIVT

participants had more frequent use of *will*, and the LIVT participants had a rather strong tendency to use *it* in the collocation *it will* in their English.

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
WILL	209	100	**7.24**	168	100	**5.66**	111	100	**4.76**
IT WILL	27	12.92	**0.94**	11	6.55	**0.37**	5	4.50	**0.21**
L1 collocates of *IT WILL*	**THINK 6** WILL 4 YEAH 2 MAYBE 2 KNOW 2 IT 2 AND 2			EH 3 AND 3 SOMETIMES 2 YES 1 THINK 1			IT 3 EH 2		
R1 collocates of *IT WILL*	**BE 10** IT 3 EH 3 MAKE 2			**BE 5** TALK 2 IT 2			**BE 2** HELP 1 SOMETIMES 1 WORTHY 1		

Table 4.75 Use of *will* and *it will* in LIVT, TESOL and NTESOL

When taking a close look at the use of *it will* in terms of its collocates, it is found that some of the collocates are shared by all the three groups. The LIVT, TESOL and NTSOL participants all had the collocation *it will be* as their primary use of *it will*. The frequent use of *it will be* seems to be universally common, because the collocation *it will be* is also the most common use of *it will* in the BASEah and COCAsp corpora. (See Table 4.76 and Table 4.77.) However, there is one noteworthy usage in the LIVT group. It is noted that the LIVT participants had a pattern *think it will be* which was also found in the COCAsp corpus. This pattern was seldom used by the TESOL and NTESOL participants. As we have discovered in 4.2.2, the LIVT participants used the collocation *I think* very frequently in their speech. The use of *I think* is mostly used as a means to earn more time for finding appropriate words in a spontaneous and rapid speech. The LIVT participants were very likely to pick up this strategy from the English input to which they had exposed since such a strategy is frequently used by the COCAsp speakers (native speakers) in their speech. The TESOL and NTESOL participants who did not have many opportunities to receive such input therefore were not influenced/acquired this strategy as strongly as the LIVT participants were/did. From our observation on the use of *it will*, although we did not find great differences in terms of the collocations of *it will*, what we did find again is that the LIVT participants seem to behave differently in their discourse use of language, and this finding supports our hypothesis 1.

N	L1	Centre	R1
1	THINK	IT WILL	BE
2	WILL		IT
3	YEAH		EH
4	MAYBE		MAKE
5	AND		
6	IT		
7	KNOW		

Patterns of *it will* in LIVT

N	L1	Centre	R1
1	EH	IT WILL	BE
2	AND		TALK
3	SOMETIMES		IT

Patterns of *it will* in TESOL

N	L1	Centre	R1
1	IT	IT WILL	BE
2	EH		

Patterns of *it will* in NTESOL

Table 4.76 Patterns of *it will* in LIVT, TESOL and NTESOL

N	L1	Centre	R1
1	AND	IT WILL	BE
2	BUT		MAKE
3	IT		BECOME
4	THAT		NEVER
5	HOPEFULLY		WORK
6	YOU		HAVE
7	THERE		IN
8	HOPE		

Patterns of *it will* in BASEah

N	L1	Centre	R1
1	AND	IT WILL	BE
2	THINK	IT . WILL	TAKE
3	THAT	IT ? WILL	NOT
4	BUT		HAVE
5	SAYS		HELP
6	WHAT		MAKE
7	SO		COME
8	BECAUSE		GO
9	BELIEVE		GET
10	IT		HAPPEN
11	WELL		COST

Patterns of *it will* in COCAsp

Table 4.77 Patterns of *it will* in BASEah and COCAsp

4.4.3 Summary of the Use of *it*

In 4.4, we investigated the top 10 most frequent collocates of the pronoun *it* in the LIVT, TESOL and NTESOL groups. We investigated several collocates of *it* which showed noticeable differences in the frequency of use between the three groups. What we found supports our first and second hypotheses that the EFL learners with different language learning backgrounds will behave differently in their language use. The different language learning backgrounds may influence not only the linguistic aspect in their English use, but very possibly the discourse aspects of their language use as well. The findings of our investigation on the use of *it* are presented in Table 4.78.

Support for Hypothesis 1		
Differences	Possible Causes of the differences	Words
The **LIVT** participants used the words **significantly more frequently** than the **TESOL** and **NTESOL** participants.	Different discourse preferences	*it will*
The **LIVT** participants used the words **significantly more frequently** than the **NTESOL** participants.	Different discourse preferences.	*to it,*
The **LIVT** participants used the words **significantly less frequentl**y than the **NTESOL** participants.	Different discourse preferences.	*learn it*
Support for Hypothesis 2		
Differences	Possible Causes of the differences	Words
The **TESOL** participants used the words **significantly less frequently** than the **NTESOL** participants.	Input Influence	*use it*
	Different discourse preferences.	*learn it*

Table 4.78 Findings in the investigation of the use of *it*

We found several differences which support our first and second hypotheses. The findings of the use of *it will*, *to it* and *learn it* support our hypothesis 1. With regard to the collocation *it will*, we found that the LIVT participants tended to use *(I) think it will* significantly more often than the TESOL and NTESOL participants. The use of *I think* is a discourse strategy used strongly by the LIVT participants. This usage was not that strongly used by the participants in the other two groups. In the use of *to it*, we found that the LIVT participants used the collocation *to it* more frequently when talking about how their habits were formed than the NTESOL participants. They also tended to use the collocation *to it* in a fixed pattern [*get/BE-verb used to it*]. Although we can find this pattern used by the other groups, the fact that the LIVT participants used this pattern significantly more frequently

suggests that some factors might have affected the LIVT participants behaving differently from the others. However, whether the factors were relevant to the language input they received in their English environments is still unclear. With regard to the use of *learn it*, we found that the LIVT and TESOL participants showed significantly less use of *learn it* than the NTESOL participants. The LIVT and TESOL participants also tended not to use the words *learn it* in the same contexts as the NTESOL participants did. The NTESOL participants tended to associate the collocation *learn it* with a sense of reluctant obligation. This finding shows that when people use language, their attitudes will also influence how they use words as well, as what had been discussed in chapter 2 that there are many factors which influence people's language acquisition/language use. Although we cannot find the evidence to show how input influences people's language use from these examples. Nevertheless, the differences in the use of *learn it* between the groups support our hypothesis 1 and hypothesis 2 that the EFL learners with different language backgrounds will behave differently in their language use. The use of the collocation *use it* by the TESOL participants supports our hypothesis 2. The TESOL participants tended not to use the pronoun *it* after the verb *use*. This kind of use was not found very strongly in the other two groups. When we investigated the collocations of *how to use* in the TESOL group, we found that the TESOL participants were prone to use a specific noun/noun phrase after the verb *use* more strongly than the participants in the other groups. This preference in the TESOL group shows that the TESOL participants differed from the LIVT and NTESOL participants in this respect, however, further investigation is necessary in order to discover whether the TESOL participants' explicit pedagogical training plays a crucial part in it.

4.5 Conclusion

In chapter 4, we investigated the English pronouns *I, you* and *it* to test our hypotheses that:

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

We discovered several noticeable differences between the groups and the findings supporting these two hypotheses. The potential causes of the differences, as we have discussed, may not entirely be the differences of the English learning backgrounds. Differences in discourse preferences were also found to influence these participants' use of language. For instance, the LIVT participants tended to use *I just (just)* more frequently than the other participants to add an extra tone of emphasis or hedging in their speech. The LIVT participants also tended to use *I think* more frequently than the other groups. The use of *I think* is a discourse strategy to earn more time to generate the appropriate words for their sentences. Although the participants in the other two groups also used this expression, their tendency was not as strong as the LIVT participants. We also found evidence for our hypothesis 1 which indicated that the differences might be caused by different English input. The use of *like* before *if you* in the LIVT group showed a great similarity to the use found in the reference corpora. In spoken English, the word *like* in this situation often functions in a manner similar to a connective linking ideas or a filler that helps the speaker find time for their next words. The LIVT participants were primed to use the word *like* in a similar way, and the TESOL and NTESOL participants did not have such a tendency. Therefore, it is very likely that this difference between the LIVT participants and the participants in the other two groups was caused by their different English input.

Support for Hypothesis 1	
Possible Causes of the differences	
Different discourse preferences	Input Influence
I just/ just; I think/ think; I will/ will yeah and I/yes and I	*yeah I/ yeah; yes I/yes; I would/ would; the use of past-tense verbs*
when you	*what you; (like)if you; you to*
it will, to it, learn it	
Support for Hypothesis 2	
Possible Causes of the differences	
Different discourse preferences	Input Influence
	think that/that; the use of past-tense verbs
if you	*when you*
learn it	*use it*

Table 4.79 Findings in the investigation of the use of pronouns *I*, *you* and *it*

We also discovered some evidences for hypothesis 2. Some differences were caused by the different discourse preferences. The TESOL participants used *if you* significantly more often than the NTESOL participants. We did not however find great difference in terms of its collocation and colligation (grammatical features). The finding suggested, though, that the TESOL participants tended to use more *if*-conditional clauses than the NTESOL participants. This amounts to a different discourse preference between these two groups. We found that the TESOL participants tended to avoid using the pronoun *it* when they used the verb *use*. They tended to spell out the item in full. This special tendency might be influenced by their English training, because the training and input which the TESOL participants were likely to have received would have been in a relatively more academic and written style. The precise indication of things is very likely a part of emphases. However, we cannot be sure of this. If one wants to find out whether this is really the cause, a further investigation on their training and input will be necessary. Some of these findings are potentially compatible with what we have discussed in chapter 2. Different language input will prime speakers to behave differently in their language use. Different kinds of input can prime the speakers to use a new collocation/colligation or to enforce/reduce the strength of using certain collocations/colligations. What we had found, to some degree, shows this kind of influences/enforcements/reductions exist.

These groups have been found to differ in their use of English. The question then arises of whether they would differ in their Chinese. In the next chapter, we will investigate the use of Mandarin pronouns 我 [I], 你 [*you*] and 它 [*it*] for our null hypotheses on the use of Mandarin Chinese.

Chapter 5 Use of 我(wǒ) [I], 你(nǐ) [you] and 它(tā) [it] by the three Taiwanese groups

5.1 Introduction

In this chapter, we will investigate the use of Mandarin pronouns 我 [I] , 你 [you]and 它 [it] in the three Taiwanese groups. The LIVT participates, who had been studying and staying in the UK (Liverpool) have been shown in the previous chapter to have different English use from the participants in the TESOL and NTESOL groups. The TESOL participants, who had not been exposed to abundant authentic English in the same way as the LIVT participants had but had undertaken TESOL courses in which they had received formal and mostly written English, were also shown to differ in their English use from the NTESOL participants. The NTESOL participants, who had not been exposed to as much authentic English as the LIVT participants had and had not received as much formal/academic English training as the TESOL participants had, were likewise shown in the previous chapter to differ in their English use from both the LIVT and the TESOL participants. In this chapter, we intend to examine whether there are any noticeable differences in the Mandarin spoken by the three groups. The LIVT, TESOL and NTESOL participants are native Mandarin speakers who were born and raised up in Taiwan. They all had received similar Mandarin input and their educational backgrounds were alike. In theory, therefore, their Mandarin use should differ little. Our null hypotheses, given in chapter 2 and repeated here for convenience, are therefore:

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

If either of the null hypotheses were found to be incorrect, the implication would be that the exposure to new English environments (input) or specific training of English may potentially be the factors of influences.

In this chapter, in parallel to the investigation of the use of English pronouns by the three groups, we will examine the collocations of the pronouns 我 [I] , 你 [you]and 它 [it] in their spoken Mandarin.

5.2 The use of 我 (wǒ) [I]

5.2.1 L1 Collocates of 我 (wǒ) [I]

In the list of the top 10 most frequent L1 collocates of the Mandarin first person pronoun 我 [I] (Table 5.1), the three groups are found to have similar collocates. They all used the connectives 因为 [because], 所以 [so/therefore], 可是 [but] as L1 collocates of 我 [I] very frequently. The discourse marker 对 [yes/yeah] is also found as a L1 collocate with a high frequency in all of the three groups' lists. The interjection sounds 啊 [ah], 呃 [uh] and 嗯 [eh] and the repetition of the pronoun 我 [I] occur frequently as well.

	LIVT	1711	44.36	TESOL	1270	35.96	NTESOL	1585	37.63
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	因为	113	2.93	因为	83	2.35	因为	110	2.61
2	所以	94	2.44	对	63	1.78	所以	101	2.40
3	我	76	1.97	然后	61	1.73	对	79	1.88
4	对	70	1.81	所以	61	1.73	嗯	69	1.64
5	就是	64	1.66	嗯	45	1.27	**觉得**	50	1.19
6	啊	62	1.61	就是	39	1.10	然后	49	1.16
7	可是	57	1.48	可是	39	1.10	就是	47	1.12
8	嗯	54	1.40	但是	32	0.91	我	46	1.09
9	**觉得**	54	1.40	**是**	29	0.82	但是	44	1.04
10	**是**	47	1.22	我	27	0.76	那	38	0.90
--	但是	113	2.93	**觉得**	**25**	**0.71**	**是**	**30**	**0.71**

Table 5.1 L1 Collocates of 我(wǒ) in LIVT, TESOL and NTESOL

In Table 5.1, we can see that the LIVT participants used the interjection 啊 [ah] in front of 我 [I] and the repetition of 我 [I] much more frequently than the TESOL and NTESOL participants. The frequent use of collocations 啊 我 [ah I] and 我 我 [I I] in the LIVT group even shows

statistically significance in comparison with the one in the other two groups. (See Table 5.2.) This does not suggest that the LIVT participants were prone differently in using the word 我 [I]. But it does show that the LIVT participants' performance in speech had a special feature. The frequent use of the collocations 啊 我 [ah I] and 我 我 [I I] suggests that the LIVT participants were likely to hesitate or speak without having a ready sentence more frequently than the TESOL and NTESOL participants when talking about themselves. This finding of a high use of the collocations 啊 我 [ah I] and 我 我 [I I] in the LIVT group does not agree with our null hypothesis 3, in which we hypothesised that there would be no noticeable differences in the use of Mandarin Chinese between the Taiwanese EFL learners studying in the UK and the Taiwanese EFL learners studying in Taiwan. This difference between the groups was probably a difference resulting from the participants' different discourse behaviour. We find the native Mandarin speakers with different English language learning backgrounds had statistically significant differences between each other. This finding challenges our null hypothesis. It suggests that there are some interesting aspects to be investigated further. We start by considering the use of 觉 得 [to think] .

LIVT - NTE						**LIVT - TE**						**TE - NTE**					
Over						Over						Over					
Words	With	LLR	Sig.			Words	With	LLR	Sig.			Words	With	LLR	Sig.		
啊	我	11.140	0.001	***	+	我	我	20.948	0.000	***	+	然后	我	5.552	0.018	*	+
我	我	10.239	0.001	**	+	啊	我	9.041	0.003	**	+	Less					
可是	我	10.106	0.001	**	+	**觉得**	我	**8.547**	**0.003**	****	**+**	Words	With	LLR	Sig.		
是	我	**4.965**	**0.026**	****	**+**	所以	我	4.525	0.033	*	+	所以	我	6.974	0.008	**	-
就是	我	**3.899**	**0.048**	****	**+**	**就是**	我	**4.523**	**0.033**	****	**+**						

Table 5.2 Log-likelihood Ratios for L1 collocates of 我(wǒ) in LIVT, TESOL and NTESOL

(我(wǒ) 觉得(juéde) 我(wǒ) [I think I]

The verb 觉得(jué de) [to think] is found as a L1 collocate of 我(wǒ) [I], which occurrs only in the LIVT and NTESOL groups' top 10 most frequent L1 collocate lists. The TESOL participants did not use 觉得我(jué de wǒ) [think I] as frequently as the participants in the other two groups. However, when checking the use of 觉得我(jué de wǒ) [think I] via log-likelihood ratio across the groups, it is found that the TESOL participants used 觉得我(jué de wǒ) [think I] significantly less often than the LIVT participants. There are 25 occurrences of 觉得我(jué de wǒ) [think I] found in the TESOL data (0.71 times per 1000 words.) The use of 觉得我(jué de wǒ) [think I] in the LIVT group is 1.42 times per 1000 words, and 1.21 times per 1000 words in the NTESOL group. The LIVT participants used this collocation 觉得我(jué de wǒ) [think I] twice as often as the TESOL participants.

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
觉得 (juéde)	685	100	17.76	559	100	15.83	781	100.00	18.54
我(wǒ) 觉得 (juéde)	487	71.09	**12.63**	372	66.55	**10.53**	519	66.45	**12.32**
(我(wǒ)) 觉得 (juéde) 我(wǒ)*	54	7.88	1.40	25	4.47	0.71	50	6.40	1.19
think	599	100	20.75	478	100	16.10	406	100	17.42
I think	507	84.64	**17.56**	425	88.91	**14.32**	354	87.19	**15.19**
I think I	34	5.68	1.18	30	6.28	1.01	26	6.40	1.12

Table 5.3 Use of 我(wǒ) 觉得(juéde) and *I think* in LIVT, TESOL and NTESOL

When we observe how the collocation 觉得我(jué de wǒ) [think I] was used by these three groups, it is found that this collocation occurrs mostly in the pattern 我觉得我(wǒ jué de wǒ) [I think I] in the LIVT group. In the total 33 instances of 觉得我(jué de wǒ) [think I] found in the LIVT group, there are 29 occurrences of 我觉得我(wǒ jué de wǒ) [I think I] and the other 4 instances are variants of 我觉得我(wǒ jué de wǒ) [I think I] (我(wǒ) * 觉得我(jué de wǒ) [I think I]). There are 2 instances of 我自己觉得我(wǒ zì jǐ jué de wǒ) [I myself think

I], 1 instance of 我 会 觉 得 我 [I will think I] and 1 instance of 我 是 觉 得 我 [I + (emphasis marker SHI) + think I]. The TESOL participants did not show this usage very frequently. This finding is similar to what we have found with regard to the use of *I think* in chapter 4.2.2. Previously in the use of *I think*, we have discovered that the LIVT participants used the collocation *I think* significantly more frequently than the TESOL and NTESOL participants. The use of *I think* is a discourse strategy to start a sentence. The use of (我) 觉 得 (我) [I think I] in Mandarin Chinese is similar to the use of *I think*. In the later investigation on the use of R1 collocates of 我 [I] (5.2.2), we will see that 我 觉 得 [I think] is the most frequent collocation in the three groups. This finding suggests that the use of 我 觉 得 [I think] as a discourse starter is commonplace to all the three groups.

The tendency to use this kind of discourse usages more often or less often is consistent in the LIVT and TESOL groups in their spoken Mandarin and spoken English. The LIVT participants used both *I think (I)* and (我) 觉 得 (我) [I think I] more often than the participants in the other two groups, and the TESOL participants used both *I think (I)* and (我) 觉 得 (我) [I think I] less often than the participants in the other two groups. This finding suggests that it is a discourse strategy used in their speech regardless of which language they used. However, the NTESOL participants showed an inconsistent tendency with regard to using this discourse strategy in their spoken English and spoken Mandarin. The NTESOL participants used (我) 觉 得 (我) [I think I] as often as the LIVT participants in Mandarin, but used *I think (I)* less often than the LIVT participants in English. (See Table 5.3.)

This finding suggests that there might be an influence from one language to the other depending on how familiar the participants were in the use of both languages. This finding may potentially be compatible with lexical priming theory discussed in chapter 2 but with a twist. The LIVT participants used the collocation (我) 觉 得 (我) [I think I] as a discourse strategy in Mandarin frequently. When they learned the collocation *I think* which can serve for a similar discourse purpose in English (and received and produced this collocation frequently), the priming in their minds for the use of the collocation *I think* in English and the collocation (我) 觉 得 (我) [I think I] in Mandarin might be conjunct together. Therefore, there were little differences in such discourse use in the LIVT participants' spoken English and spoken Mandarin. However, the NTESOL participants, who also used the collocation (我) 觉 得 我 [I think I] as a discourse strategy in their Mandarin frequently, had not been primed to use the English collocation *I think* for the same

discourse purpose as strongly as its Mandarin equivalent. The reason behind this lack of use is very likely that the NTESOL participants had only been primed weakly in their use of *I think* due to the lack of exposure to English input or of opportunities to use it. Therefore, it lead to the inconsistency that we found in the NTESOL participants' use of spoken English and spoken Mandarin Chinese.

In Table 5.2, it is also found that the LIVT participants used the collocation 可 是 我 [but I] significantly more frequently than the NTESOL participants. The TESOL participants likewise used the collocation 然 后 我 [then I] significantly more frequently than the NTESOL participants, and the TESOL participants used the collocation 所 以 我 [so I] significantly less frequently than the LIVT and NTESOL participants. We hypothesised that the LIVT participants, who are native Mandarin speakers with one year studying abroad experience, and the TESOL participants, who are native Mandarin speakers having explicit pedagogical English training, will show no noticeable differences in their use of Mandarin Chinese in comparison with the NTESOL participants. The findings of these noticeable differences between groups challenge our null hypotheses. These differences may represent the different discourse preferences between groups. The findings of these preferences for using the connectives 可 是[but] , 所以[so, therefore] and 然 后[then] with 我 [I] suggest that the participants in question might be relying on certain connectives to organise their ideas more frequently than the other groups of participants. Although it would be worth studying how the participants with different language learning backgrounds organise their ideas in speaking, this is beyond the concerns of the current study. Therefore, we will not further investigate the use of these collocations.

It is found, though, that there are two collocations in the LIVT group that stood out and may be regarded as important. The first one of these collocations is 是 我 [SHI (to be) + I], and the second is 就 是 我 [just I]. The LIVT participants used the collocation 是 我 [SHI (to be) + I] much more frequently than the NTESOL participants. The word 是 in Mandarin means 'to be'. It often serves a similar use to BE-verbs in English-Chinese translation, as shown in the sample below.

我	是	学 生。	= I am a student.
wǒ	shì	xuésheng	
I	be	student	

However, the word 是 functions in wider range of roles than the full BE-verbs in English. It can also be a formal discourse marker meaning 'yes', and can serve as a noun meaning 'correctness'. It is often used as a morpheme in some connectives and adverbs, such as the connectives 但 是 [but],

可 是 [but] and the adverb 就 是 [just]. The collocation 是 我 here excludes the use of 是 as a morpheme. But even so, the LIVT participants used 是 我 in their speech more often than the NTESOL participants. It suggests that there may be some use of 是 我 that differentiated the LIVT participants from the NTESOL participants. Because of the complexity of the use of 是, the use of 是 earns its own chapter. We will therefore postpone discussion of the anomalous use of 是 until chapter 6. Here we will focus on the second collocation 就 是 我 [just I].

The collocation 就 是 我 [just I] was used by the LIVT participants more frequently than by the TESOL and NTESOL participants. The phrase 就 是 [just I] is similar in meaning and function to the adverb *just* in English. In the case of the collocation 就 是 我 [just I], it is used to add emphasis and modality into the speech. Our null hypothesis is that there would be little differences between the use of words in the participants of the three groups in terms of the collocations and colligations (grammatical features) of 我 [I]. However, the statistically significant high occurrences of the collocation 就 是 我 in the LIVT group (see Table 5.2) as opposed to the TESOL and NTESOL groups suggest that the collocations may differ from group to group, even though the participants were all Mandarin speakers from the same region and were asked the same questions in interviews. Therefore, it is worth having a closer look at the use of the collocation 就 是 我 and the phrase 就 是.

就是(jiùshì) 我 (wǒ)[just I]

Our initial assumption was that the LIVT participants were using the adverbial phrase 就 是 [just] more frequently in general, therefore lead to a higher use of 就 是 我 [just I]. However, when investigating the total frequency of the phrase 就 是 [just], it is found that the LIVT participants used 就 是 [just] in similar frequency as the TESOL and NTESOL participants. (See Table 5.4.)

The LIVT participants used the phrase 就是 [just] 27.60 times per 1000 words. The TESOL participants used 就是 [just] 26.34 times per 1000 words, and the NTESOL percipients used it 29 times per 1000 words. These standardised occurrences of 就是 [just] indicate that the LIVT participants did not differ too much from the TESOL and NTESOL participants in terms of frequency. However, the LIVT participants showed a tendency to use 就是 [just] in front of the person pronoun 我 [I] slightly more often than the TESOL and NTESOL participants did. Interestingly, when checking how the collocations 就是 我 [just I] was used, it is found that it was used in similar patterns with similar collocates by these three groups. The collocation 就是 我 [just I] appears in the pattern 就是 我 觉得 [just I think] in all the three groups. (See Table 5.4.)

The collocation/pattern 就是 我 觉得 [just I think] is a discourse strategy that provides the speaker with time to generate the sentence that follows. We can see that all the participants in the three groups had this use as their primary use of the collocation 就是 我. However, it is the LIVT participants who used 就是 我 [just I] most strongly in Mandarin in comparison with the other participants. These findings indicate that, similarly to what we found in the English use of *I just* and *I think* in chapter 4, the LIVT participants tended to use word/word sequences which serve a discourse function more often than the participants in the other two groups. It is difficult to know for sure why the LIVT participants would have this kind of discourse preference (using phrases which can prolong the preparation time for getting to the main idea) particularly more often than the other participants in both English and Mandarin. What we have found, however, is that the LIVT participants seemed to be more consistent in using certain combinations of words for the same discourse purpose in both English and Mandarin.

	LIVT			TE			NTE		
	Freq.		Per 1000 words	Freq.		Per 1000 words	Freq.		Per 1000 words
就是 (jiùshì)	1046		**27.12**	916		**25.94**	1214		**28.82**
就是 (jiùshì) 我 (wǒ)	64 [4]		1.66	39		1.10	47		1.12
R1 Collocates of 就是 (jiùshì) (TOP 5)	R1	Freq.	Per 1000 words	R1	Freq.	Per 1000 words	R1	Freq	Per 1000 words
	你 (nǐ)	70	1.81	**我** (wǒ)	39	1.10	我 (wǒ)	48	1.14
	我 (wǒ)	65	1.69	会 (huì)	37	1.05	它	39	0.93
	那 (nà)	32	0.83	你 (nǐ)	36	1.02	你 (nǐ)	38	0.90
	可能 (kě néng)	28	0.73	就是 (jiùshì)	23	0.65	就是 (jiùshì)	31	0.74
	比较 (bǐ jiào)	27	0.70	它 (tā)	23	0.65	有 (yǒu)	30	0.71
Patterns of *就是 (jiùshì) 我 (wǒ)*	N / L1 / Centre / R1 1 / 就是 / 就是我 / 觉得 / jiùshì / jiùshìwǒ / juédé / just / just I / think 2 / / / 没有 / / / méiyǒu / / / not have			N / L1 / Centre / R1 1 / 就是 / 就是我 / 觉得 / jiùshì / jiùshìwǒ / juédé / just / just I / think			N / L1 / Centre / R1 1 / / 就是我 / 觉得 / / jiùshìwǒ / juédé / / just I / think		

Table 5.4 The use of 就是 (jiùshì) and 就是 (jiùshì) 我 (wǒ) in LIVT, TESOL and NTESOL

[4] WordSmith Tools 6.0 sometimes showed this kind of inconsistency in calculating the occurrences of collocates. In this case, when using the word 我 (wǒ) [I] as the node, there were 64 occurrences of 就 (jiù) as its L1 collocate. However, when we used the phrase 就是 (jiùshì) as the node, there were 65 occurrences of the word 我 (wǒ) [I] as its R1 collocate. The accurate count of occurrences of the collocation 就是 (jiùshì) 我 (wǒ) (including 就是 (jiùshì), 我 (wǒ)) is in fact 65 occurrences. This kind of inconsistency happened frequently when analysing Mandarin texts. The figures did not differ greatly and had little effect when we use them to calculate the log-likelihood ratios. However, we still had to point out that such an inconsistency of figures happened.

5.2.2 R1 Collocates of 我 (wǒ)

In the top 10 most frequent R1 collocates of the Mandarin pronoun 我 [I], we can see there are plenty of shared collocates in the three groups. The most frequent collocate in the three groups is the verb 觉得 [think]. It is similar to the use of *I think* discussed previously. The collocation of 我觉得 [I think] serves a discourse function and is usually found in initial position in a sentence. The R1 collocates, such as 就 [just], 会 [suggesting capability or the possibility of achieving something], 不 [no/not], 是 [to be], 也 [also], 的 [possessive marker], are found in the top 10 most frequent R1 collocate lists of the three groups. There are, however, some R1 collocates that appear only in one or two groups' lists, such as 还 [still, yet], 没(有) [not have, not to be], 现在 [now; at the moment] and 自己 [oneself]. (See Table 5.5.)

	LIVT	1711	100	44.36	TESOL	1270	100	35.96	NTESOL	1585	100	37.63
N	R1	Freq.	%	per 1000 words	R1	Freq.	%	per 1000 words	R1	Freq.	%	per 1000 words
1	觉得	482	28.17	12.50	觉得	372	29.29	10.53	觉得	519	32.74	12.32
2	就	111	6.49	2.88	就	86	6.77	2.44	会	82	5.17	1.95
3	不	98	5.73	2.54	不	61	4.80	1.73	不	61	3.85	1.45
4	我	77	4.50	2.00	会	55	4.33	1.56	也	56	3.53	1.33
5	会	66	3.86	1.71	是	37	2.91	1.05	就	48	3.03	1.14
6	是	42	2.45	1.09	的	32	2.52	0.91	我	47	2.97	1.12
7	也	41	2.40	1.06	也	32	2.52	0.91	自己	36	2.27	0.85
8	的	33	1.93	0.86	我	28	2.20	0.79	的	30	1.89	0.71
9	没	27	1.58	0.70	没	20	1.57	0.57	还	26	1.64	0.62
10	现在	25	1.46	0.65	自己	18	1.42	0.51	是	20	1.26	0.47

Table 5.5 R1 Collocates of 我(wǒ) in LIVT, TESOL and NTESOL

LIVT - NTE						LIVT - TE						TE - NTE					
Over						Over						Over					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
我	**就**	31.569	**0.000**	***	+	我	我	**19.074**	0.000	***	+	**我**	**就**	18.598	**0.000**	***	+
我	不	15.651	0.000	***	+	**我**	**觉得**	**7.129**	**0.008**	**	+	**我**	**是**	8.558	**0.003**	**	+
我	**是**	10.055	**0.002**	**	+							Less					
我	我	9.720	0.002	**	+							With	Word	LLR	Sig.		
我	现在	8.043	0.005	**	+							**我**	**还**	12.146	**0.000**	***	-
Less												**我**	**觉得**	5.455	**0.020**	*	-
With	Word	LLR	Sig.														
我	自己	5.382	0.020	*	-												
我	**还**	4.113	**0.043**	*	-												

Table 5.6 Log-likelihood Ratios for R1 collocates of 我(wǒ) in LIVT, TESOL and NTESOL

When checking these collocations with log-likelihood ratios (See Table 5.6), it is found that the LIVT and TESOL participants used the collocations 我 就 [I just] and 我 是 [I to be] significantly more often in comparison to the NTESOL participants. The LIVT and TESOL participants used the collocations 我 还 [I still/yet] significantly less often than the NTESOL participants. The LIVT participants used the collocation 我 觉得 [I think] significantly more often than the TESOL participants, and the collocations 我 不 [I not] and 我 现在 [I now] significantly more often than the NTESOL participants. The LIVT participants used the collocation 我 自己 [I myself] significantly less often than the NTESOL participants. The TESOL participants used 我 觉得 [I think] significantly less often than the LIVT and NTESOL participants. All of these significant higher or lower use of the collocation challenge our null hypotheses. The use of the collocation 我是[I BE], like the earlier collocation 是 我[BE I], will be discussed in chapter 5. The use of 我 就 [I just] is similar to the use of *I just* in English. As mentioned previously, the LIVT participants used *I just* and the word *just* significantly more often than the TESOL and NTESOL participants. We found that the use of *just* is for the discourse purpose of showing modesty or emphasis in their speech. Here we can see that the LIVT participants showed consistency in using *(I) just* and 我 就 [I just] for this discourse purpose in their English and Mandarin. However, here we can also see that the TESOL participants used 我 就 [I just] in their Mandarin as often as the LIVT participants, but in English, the TESOL participants did not use *I just* that often. We will further investigate the use of 我 就 [I

just] later. We will also investigate the use of 我 还 [I still], since in Mandarin the word 还 is often used as an adverb to express a toning down (meaning 'fairly/rather'). We will want to see, beside the significantly high use of 我 还 in the LIVT and TESOL groups, whether there are any other noteworthy findings that might suggest that the LIVT and TESOL participants were using the words differently as regards their discourse purposes from the NTESOL participants. The high use of 我 不 [I not] and 我 现在 [I now] in the LIVT group in comparison with the NTESOL group suggests that the LIVT participants might use negation more often than the NTESOL participants and might identify the time of happening with time adverbs more often than the NTESOL participants. But although it would be interesting to see how these participants used negation or used time adverbs to indicate the time (sequence) in detail, space prohibits investigation of these two topics. A similar situation happens with regard to the use of 我 自己 [I myself]. The use of 我 自己 [I myself] contains a discourse function of adding emphasis when the speaker is referring to him/herself. 我 自己 [I myself] carries a sense of 'it is I and I alone'. The finding that the NTESOL participants used this expression more often than the LIVT participants suggests that the NTESOL participants may have a stronger need to express the individuality of themselves. An explanation of this special tendency is not easy to find. To find the answer, a thorough discourse analysis would almost certainly be needed, which goes beyond the scope of this study. Therefore, we will not further investigate the use of 我 自己 [I myself]. Instead, we will investigate the uses of the collocations 我 就 [I just] and 我 还 [I still].

我 就 (wǒ jiù)[I just]

The collocation 我 就 [I just] was used significantly more frequently by the LIVT and TESOL participants than by the NTESOL participants. In fact, the word 就 on its own was also used significantly more frequently by the LIVT and TESOL participants. (See Table 5.7.)

我 就 (wǒ jiù)	LLR	Sig.		
LIVT - TE	1.393	0.238		+
LIVT - NTE	31.569	**0.000**	***	+
TE - NTE	18.598	**0.000**	***	+
就 (jiù)	LLR	Sig.		
LIVT - TE	0.82	0.365		+
LIVT - NTE	44.06	**0.000**	***	+
TE - NTE	30.93	**0.000**	***	+

Table 5.7 Log-likelihood Ratios: 我就 (wǒ jiù) and 就 (jiù) in LIVT, TESOL and NTESOL

The Mandarin word 就 has many functions. It can be a verb which means 'to approach' or 'to engage' and a preposition which means 'with regard to' (usually in written Mandarin). The primary function of 就 is that of adverb, and it can mean 'right away', 'only', 'already' and 'just' with a tone of emphasis. The high frequencies of the use of 就 and the collocation 我就 in the LIVT and TESOL groups suggest that the LIVT and TESOL participants tended to modify and emphasise what they were saying more often than the NTESOL participants. Especially when the LIVT and TESOL participants expressed an idea relating their personal affairs, they tended to use the word 就 to modify their tone more often than the NTESOL participants. When taking a closer look at the collocates of 就 and 我就 it is found that these two words/collocations did not differ greatly. (See Table 5.8.)

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
就 (jiù)	796	20.64	829	23.48	746	17.71
我就 (wǒ jiù)	111	**2.88**	86	**2.44**	48	1.14

L1 Collocates of 就 (jiù) (TOP 5)	L1	Freq	Per 1000 words	L1	Freq	Per 1000 words	L1	Freq	Per 1000 words
	我	**111**	**2.88**	我	**86**	**2.44**	然后	50	1.19
	然后	54	1.40	然后	56	1.59	**我**	**48**	**1.14**
	你	45	1.17	你	37	1.05	你	36	0.85
	可能	39	1.01	啊	28	0.79	话	31	0.74
	啊	33	0.86	的	22	0.62	可能	28	0.66

Patterns of 我就 (wǒ jiù) (TOP 5)	N	L1	Centre	R1
	1	然后	我就	觉得
		ránhòu	wǒjiù	juédé
		then	I just	think
	2	所以		不会
		suóyǐ		búhuì
		so		cannot
	3	时候		没有
		shíhou		méiyǒu
		time/period		not have
	4	东西		可以
		dōngxi		kéyǐ
		thing		can
	5	不错		喜欢
		búcuò		xǐhuan
		not bad		like

	N	L1	Centre	R1
	1	然后	我就	觉得
		ránhòu	wǒjiù	juédé
		then	I just	think
	2	所以		赶快
		suóyǐ		gǎnkuài
		so		hurry
	3	这样		不要
		zhèyàng		búyào
		this way		not want
	4	知道		
		zhīdào		
		know		
	5	小说		
		xiǎoshuō		
		fiction		

	N	L1	Centre	R1
	1	然后	我就	觉得
		ránhòu	wǒjiù	juédé
		then	I just	think
	2	所以		赶快
		suóyǐ		gǎnkuài
		so		hurry
	3	时候		没有
		shíhou		méiyǒu
		time/period		not have

Table 5.8 Collocates and patterns of use of 我就(wǒ jiù) and 就(jiù) in LIVT, TESOL and NTESOL

The use of 就[just] as an adverb in all three groups shows high similarity in its collocates. The words 然后 [then], 我 [I], 你 [you] and 可能 [maybe] are the most frequent L1 collocates in the three groups of 就[just]. It is found that the collocation 我就 is used in the pattern 然后 我 就 觉得 [then I just think] very frequently in all the three groups. Investigation of the collocates of the word 就 [just] and the combination 我就[I just] shows that the participants in the three groups did not use 就 [just] and 我 就 [I just] very differently in terms of their collocations. It is their different frequencies of use of the word 就 [just] as a discourse strategy that resulted in the significant differences between the LIVT and TESOL participants and the NTESOL participants.

I just	LLR	Sig.			我 就 (wǒ jiù)	LLR	Sig.		
LIVT - TE	10.359	**0.001**	**	+	LIVT - TE	1.393	0.238		+
LIVT - NTE	6.952	**0.008**	**	+	**LIVT - NTE**	31.569	**0.000**	***	+
TE - NTE	0.142	0.706		-	**TE - NTE**	18.598	**0.000**	***	+
just	LLR	Sig.			就 (jiù)	LLR	Sig.		
LIVT - TE	54.61	**0.000**	***	+	LIVT - TE	0.82	0.365		+
LIVT - NTE	10.34	**0.001**	**	+	**LIVT - NTE**	44.06	**0.000**	***	+
TE - NTE	14.21	**0.000**	***	-	**TE - NTE**	30.93	**0.000**	***	+

Table 5.9 Log-Likelihood Ratios: *I just, just,* 我就 (wǒ jiù) and 就 (jiù)

	LIVT		**TE**		**NTE**	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
就 (jiù)	796	20.64	829	23.48	746	17.71
我就 (wǒ jiù)	111	**2.88**	86	**2.44**	48	**1.14**
just	321	11.12	166	5.59	194	8.32
I just	67	**2.32**	36	**1.21**	31	**1.33**

Table 5.10 Log-Likelihood Ratios: *I just, just,* 我 就 (wǒ jiù) and 就 (jiù)

However, similarly to what we found with regard to the inconsistency in the frequencies of the use of *I think* and 我觉得 [I think] within the same group, it is also found this inconsistency in the use of *I just* and 我就 [I just]. The LIVT participants showed a consistency in thier use of the words *I just (just)* and 我就 (就) as a discourse strategy in both English and Mandarin, and they used these expressions comparatively more often than the other participants. The NTESOL participants used the words *I just (just)* and 我就 (就) [I just (just)] as a discourse strategy comparatively less often than the other two groups. However, the TESOL participants showed an inconsistency in the use of their English and Mandarin. (See Table 5.9 and Table 5.10.) The TESOL participants used 我就 (就) [I just (just)] as often as the LIVT participants in Mandarin. However, the TESOL participants did not use the similar expression(s) *I just* (*just*) as often as the LIVT participants. This finding suggests that there might be an influence of one language on the other depending on how familiar the participants are with both languages. This finding may be compatible with lexical priming theory again with a twist. The LIVT participants used the collocation 我就 (就) (I just (just)) as a discourse strategy in Mandarin frequently. When they learned the collocation *I just*

(just) which can serve a similar discourse purpose, they might be primed to use it frequently in their English. Their priming in the use of 我 就 (I just) and *I just* showed no great differences in terms of the discourse use. However, the TESOL participants, who also used the collocation 我 就 (就) [I just] as a discourse strategy in their Mandarin frequently, had not been primed to use the English collocation *I just (just)* for the same discourse purpose as strongly as they had been for the Mandarin expression. It is possible that the TESOL participants had been primed weakly in their use of *I just*. However, unlike the use of *I think* in the NTESOL group that we discussed previously, the reason behind this weak priming for *I just* (*just*) in the TESOL group is probably not lack of exposure to English input, but lack of exposure to casual spoken English input in which the use of *just* is more likely to be heard. Nevertheless, what we have found is that the LIVT participants showed a stable consistency in their use of certain words and phrases which can function for a similar discourse purpose both in their spoken English and spoken Mandarin. This finding suggests that there is a possibility that for the LIVT participants, who had been exposed to comparatively abundant English input, the division between their primings for certain English words/phrases and Mandarin words/phrases, which shared some similarity in use, became less distinct. The finding suggests that there is a possibility that there is cross language interference in people's primings.

我 还 (wǒ hái) [I still/fairly]

With regard to frequencies of the collocation 我 还 [I still/fairly] in the three groups, the NTESOL participants used the collocation 我 还 [I still/fairly] significantly more often than the LIVT and TESOL participants. However, when checking the overall occurrences of the word 还 on its own, it is found that the participants in the three groups used it with similar frequencies. (See Table 5.11.)

The Mandarin word 还 has many meanings. It can be used as a verb meaning 'to return' (pronounced in *huán*), but its primary use is as an adverb meaning 'still', 'yet' or 'fairly' (pronounced in *hái*). In its occurrences, the word 还 is used mostly as an adverb in the three groups.

我 还 (wǒ hái)	LLR	Sig.		
LIVT - TE	2.398	0.121		+
LIVT - NTE	4.113	**0.043**	*	-
TE - NTE	12.146	**0.000**	***	-
还 (hái)	LLR	Sig.		
LIVT - TE	3.56	0.059		-
LIVT - NTE	6.41	**0.011**	*	-
TE - NTE	0.31	0.580		-

Table 5.11 Log-likelihood Ratios for 我 还 (wǒ hái) and 还 (hái) in LIVT, TESOL and NTESOL

In Table 5.12, moreover, we can find that the NTESOL participants used this adverb 还 [still/fairly] with 我 [I] in their speech more often in comparison to the LIVT and TESOL participants. The collocation 我 还 [I still] occurrs 0.62 times per 1000 words in the NTESOL group, while it occurrs 0.31 times per 1000 words in the LIVT group and merely 0.14 times per 1000 words in the TESOL group. However, these different frequencies of the use of 我 还 does not appear to show differences in the use of 我 还. When checking the collocations for the word 还 and the collocation 我还, as we found in respect of 我 就 [I just], their collocates does not differ very greatly. Especially with regard to the use of the collocation 我 还, it is found that the collocation 我 还 appears in the pattern 我 还 蛮 喜 欢 [I fairly/quite like] very frequently. (The word 蛮 is often used as an adverb meaning 'quite' in Mandarin, and the word 还 collocates with 蛮 frequently, also meaning 'quite'.) This usage is found in all the groups. (See the instances below in Table 5.13.)

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
还 (hái)	153	3.97	173	4.90	218	5.18
我 还 (wǒ hái)	12	0.31	5	0.14	26	0.62

L1 Collocates of *还 (hái)* (TOP 5)	L1	Freq	Per 1000 words	L1	Freq	Per 1000 words	L1	Freq	Per 1000 words
	rán hòu 然后	12	0.31	jué de 觉得	16	0.45	wǒ 我	26	0.62
	wǒ 我	12	0.31	rán hòu 然后	16	0.45	rán hòu 然后	23	0.62
	jué de 觉得	10	0.26	jiù shì 就是	14	0.40	jué de 觉得	17	0.55
	jiù 就	10	0.26	shì 是	10	0.28	dōu 都	9	0.40
	dé 得	6	0.16	dōu 都	9	0.25	shì 是	9	0.21

Patterns of 我 还 (wǒ hái) (TOP 5)

N	L1	Centre	R1	R2
1	发现	我还	蛮	喜欢
	fāxiàn	wǒhái	mán	xǐhuan
	find	I still	quite	like

N	L1	Centre	R1	R2
1	所以	我还	蛮	喜欢
	suóyǐ	wǒhái	mán	xǐhuan
	so	I still	quite	like

N	L1	Centre	R1	R2
1	所以	我还	蛮	喜欢
	suóyǐ	wǒhái	mán	xǐhuan
	so	I still	quite	like

Table 5.12 Use of 我 还 (wǒ hái) and 还 (hái) in LIVT, TESOL and NTESOL

LIVT	
1	[我] 发现 **我 还 蛮 喜欢** 心理学 的
	[wǒ] fāxiàn **wǒ hái mán xǐhuan** xīnlǐxué de
	I found that I'm fairly/quite fond of psychology.
2	有 一 个 作家 **我 还 蛮 喜欢** 的
	yǒu yī gè zuòjiā **wǒ hái mán xǐhuan** de
	There is a writer whom I'm fairly/quite fond of.
3	**我 还 蛮 喜欢** 这样 类型 的 书
	wǒ hái mán xǐhuan zhèyàng lèixíng de shū
	I fairly/quite like this kind of books.
TESOL	
1	所以 **我 还 蛮** 不 **喜欢** 新诗 的
	suóyǐ **wǒ hái mán** bù **xǐhuan** xīnshī de
	So I'm not quite fond of modern poems.
2	**我 还 蛮 喜欢** 看 小说 的 啦
	wǒ hái mán xǐhuan kàn xiǎoshuō de lā
	I fairly/quite like reading a novel.
NTESOL	
1	那里 的 环境 **我 还 蛮 喜欢** 的
	nàlǐ de huánjìng **wǒ hái mán xǐhuan** de
	I fairly/quite like the environment there.
2	所以 **我 还 蛮 喜欢** 的, 嗯
	Suóyǐ **wǒ hái mán xǐhuan** de , èn
	So I fairly/quite like [it], eh.

Table 5.13 Instances of the use of 我还蛮喜欢 (wǒ hái mán xǐhuan) in LIVT, TESOL and NTESOL

The finding that the NTESOL participants used the collocation 我 还 [I still/fairly] significantly more often than the LIVT and TESOL participants but differed little in the lexico-grammatical use indicates that this is a difference of discourse preference. The use of 我 还 [I still/fairly] is a means of softening the tone of a statement, and this use may appear when a speaker wants to give a contrary thought. For instance, the pattern 我 还 蛮 喜 欢 + NP/的 [a marker used at the end of declarative sentence] contains a sense of 'This is probably unexpected, but I do like [something].' The more frequent use of 我 还 [I still/fairly] in the NTESOL group may indicate that, in comparison with the LIVT and TESOL participants, the NTESOL participants might have presupposed what kind of answers that the interviewer would like to hear. The cause of this difference in the discourse use, although it may be linked to the different language learning backgrounds, is more relevant to the discourse factor and the learners' attitudes instead of the lexico-grammatical factor.

5.2.3 Summary of the use of 我 (wǒ) [I]

In our investigation of the use of 我 (wǒ) [I], we have found several noticeable differences between the three groups. Therefore, our null hypotheses were found to be incorrect. The findings of our investigation into the use of collocations are presented in Table 5.14 and Table 5.15.

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	Collocations	The type of differences	Descriptions OR noteworthy findings/observation
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	jiù shì wǒ 就 是 我 [just I]	Discourse preference	The LIVT participants used the pattern jiù shì wǒ 就是 我 jué de 觉得 [just I think] as a discourse strategy to start their sentences more often than the TESOL and NTESOL participants. (Discourse preference)
	wǒ jiù 我 就 [I just]/ jiù 就 [just]	Discourse preference	The LIVT participants used wǒ jiù 我 就 [I just]/ jiù 就 for toning down their statements in their discourse more often than the TESOL and NTESOL participants. (Discourse preference) The LIVT participants showed consistency in using wǒ jiù 我 就 [I just] and *I just* more often than the TESOL and NTESOL participants. (Discourse preference)
The LIVT participants used the collocations **significantly less frequently** than the NTESOL participants.	wǒ hái 我 还 [I still/fairly]	Discourse preference	The LIVT participants did not use wǒ hái 我 还 [I still/fairly] to soften the statement of a contrary idea as often as the NTESOL participants. (Discourse preference)

Table 5.14 Findings in the use of 我(wǒ) that Disagree with Hypothesis 3

These noticeable differences appear to have discourse explanations. However, we found that there is a possibility that different degrees of exposure to English input may have influenced whether the participants were primed to use certain expressions (*I just*/*I think*) for particular discourse purposes more strongly or weakly.

Findings that Disagree with Hypothesis 4 (null hypothesis)			
Noticeable Differences	Collocations	The type of differences	Descriptions OR noteworthy findings/observation
The TESOL participants used the collocations **significantly less frequently** than the LIVT and NTESOL participants.	wǒ juéde wǒ 我觉得(我) [I think I]	Discourse preference	The TESOL participants showed inconsistency in using the collocation *I think* in English less often than the collocation wǒ juéde 我觉得 in Mandarin for the similar discourse purposes. The LIVT participants showed consistency in using wǒ jué dé 我觉得 [I think] and *I think* with similar frequency, and they also use the use more often than the TESOL and NTESOL participants.
The TESOL participants used the collocations **significantly less frequently** than the LIVT participants.	wǒ jiù 我就[I just]/ jiù 就(jiù) [just]	Discourse preference	The TESOL participants did not use wǒ jiù 我就[I just]/ jiù 就[just] for toning down their discourses as often as the LIVT participants. (Discourse preference) The TESOL participants showed inconsistency in using the collocation *I just* in English less often than the collocation wǒ jiù 我就[I just] in Mandarin for similar discourse purposes. (Discourse preference)
The TESOL participants used the collocations **significantly less frequently** than the NTESOL participants.	wǒ hái 我 还 [I still/fairly]	Discourse preference	The TESOL participants did not use wǒ hái 我 还 [I still/fairly] to soften the tone when giving a contrary idea in a statement as often as the NTESOL participants. (Discourse preference)

Table 5.15 Findings in the use of 我(wǒ) that Disagree with Hypothesis 4

5.3 The Use of 你[you]

In our examination of the use of the collocations of 我 [I] in the previous section, we saw that even though each group had its own tendency to use certain words more frequently or less frequently than the others, their use of the lexico-grammatical aspect (collocations/colligations) for these words did not differ greatly. It is the strength of how much they used that differed. The participants were prone to use certain words more strongly/weakly in certain contexts. We found that the participants in the three groups showed a similar use in terms of the collocations and the patterns of the words, which means, as we have pointed out in our hypotheses, these participants had received similar Mandarin input which had shaped their priming of words to be similar in lexico-grammatical level. However, what we have found in the previous section suggests that although these participants had similar lexico-grammatical primings of words, how they used these words for their discourse needs would be quite different. Next, we will investigate the use of the Mandarin second person pronoun 你[you] in these groups.

5.3.1 L1 Collocates of 你[you]

In the lists of the top 10 most frequent L1 collocates of 你 [you], it is found that the three groups share several of the same collocates. (See Table 5.16.)

	LIVT	1014	26.29	TESOL	632	17.90	NTESOL	637	15.12
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	就是	72	1.87	就是	34	0.96	如果	41	0.97
2	然后	58	1.50	然后	32	0.91	然后	38	0.90
3	因为	31	0.80	是	29	0.82	就是	38	0.90
4	你	31	0.80	那	22	0.62	那	17	0.40
5	如果	28	0.73	说	20	0.57	话	17	0.40
6	说	27	0.70	跟	18	0.51	因为	16	0.38
7	时候	27	0.70	因为	18	0.51	说	14	0.33
8	跟	25	0.65	时候	17	0.48	是	13	0.31
9	让	24	0.62	给	15	0.42	所以	13	0.31
10	是	23	0.60	话	14	0.40	你	13	0.31

Table 5.16 L1 Collocates of 你(nǐ) in LIVT, TESOL and NTESOL

The adverbs 就是 [just], 然后 [then] and the connective 因为 [because] are very frequent L1 collocates in all three groups' lists. The word 是[to be] and the verb 说 [say] are also found in the three groups' lists. There were however other words that appear only in one or two of the groups' top 10 collocate lists. The words 跟 [usually used as a preposition, similar to 'at' in English] and 时候 [period of time] appear only in the LIVT and TESOL groups' lists. The connective 如果 [if] appear only in the LIVT and NTESOL groups' lists. The word 让 [usually used as a verb, similar to 'let' or 'make' in English] only appear in the LIVT group's list, and the word 给 [usually used as a verb, means 'to give'] only appear in the TESOL group's list. Next we will check whether the differences in the frequencies of these L1 collocates have statistically significance.

LIVT - NTE						LIVT - TE						TE - NTE					
Over						Over						Over					
Word	With	LLR	Sig.			Word	With	LLR	Sig.			Word	With	LLR	Sig.		
就是	你	13.920	0.000	***	+	就是	你	10.854	0.001	***	+	是	你	9.372	0.002	**	+
时候	**你**	12.649	**0.000**	*******	**+**	你	你	9.551	0.002	**	+	跟	你	7.125	0.008	**	+
跟	你	12.403	0.000	***	+	**让**	**你**	5.844	**0.016**	*****	**+**	**时候**	**你**	5.080	**0.024**	*****	**+**
你	你	9.270	0.002	**	+	然后	你	5.547	0.019	*	+	Less					
因为	你	6.297	0.012	*	+							Word	With	LLR	Sig.		
然后	你	6.163	0.013	*	+							如果	你	9.543	0.002	**	-
说	你	5.429	0.020	*	+												
让	**你**	5.214	**0.022**	*****	**+**												

Table 5.17 Log-likelihood Ratios for L1 collocates of 你(nǐ) in LIVT, TESOL and NTESOL

In Table 5.17, we can see that the collocation 就是 你 [just you] is heavily used in the LIVT group in comparison to in the TESOL and NTESOL groups. In our previous discussion of the collocation 就是 我 [just I], we found that the LIVT participants also used the collocation 就是 我 [just I] more frequently than the TESOL and NTESOL participants. Despite of the higher frequency in the LIVT group, though, the way the LIVT participants used the collocation 就是 我 did not differ greatly from that of the TESOL and NTESOL participants. The collocation 就是 我 was used by the three groups mostly in the pattern 就是 我 觉得 [just I think] as a means to start a sentence.

In the uses of the collocation 就是 你 [just you], the situation is quite similar. The collocation 就是 你 [just you] is used as a starter of a sentence, and it is frequently used in the patterns 就是 你 要 [just BE you want] and 就是 你 可能 [just BE you can]. These usages are found in the three groups. From these findings, we can again confirm that the LIVT participants used 就是 [just]/ 就 [just] to modify their sentences much more often than the TESOL and NTESOL participants.

The LIVT participants used the collocations 你 你 [you you], 然 后 你 [(and) then you], 让 你 [let you] significantly more often than the TESOL and NTESOL participants. We will not discuss the use of repetition, and the use of the connective 然 后 [(and) then], as we have said before, falls outside the concerns of this study. We will discuss our investigation on the use of 让 你 [let you] later. The LIVT participants used the collocation 说 你 [say you] significantly more often than the NTESOL participants. The verb 说 [say] in Mandarin can be used as a verb (e.g. 说话 [say/speak]) and as a suffix to a connective (e.g. 假如说 [if/for example]). It may also be used as an emphasis marker putting extra focus on the action of saying. The use of 说 [say] will be covered in our investigation of the R1 collocate of 你[you] (see 5.3.2). Therefore, we will not discuss it further in this section. The TESOL participants used the collocation 是你 [BE you] significantly more frequently than the NTESOL participants and the collocation 如果 你 [if you] significantly less often than the NTESOL participants. The use of 是[SHI (to be)] will be discussed in chapter 6, so we will leave discussion of 是你 [to be you] to that chapter. Nor will we investigate any further the use of connective 如果 [if] with second pronoun 你 [you]. Although it would be a worthwhile topic of study to investigate how participants with different language learning backgrounds used connectives, it is impossible to cover this topic in our thesis due to limitations of space.

The LIVT and TESOL participants used the collocation 时候 你 [time you] significantly more often than the NTESOL participants. The word 时候 is often used as a noun meaning ‘time/period of time’. We will investigate the use of 时候 你 [time you] in the later section. The collocation 跟 你 [with/to you] was also used more frequently by the LIVT and TESOL participants than by the NTESOL participants. However, the participants again did not show much difference in

their use of collocations/colligations of this collocation 跟 你 [with/to you]. The noticeable difference in frequencies of the use of 跟 你 [with/to you] between the groups appeared to result from the participants' different discourse preferences.

Our null hypotheses assume that participants with different English learning backgrounds will show no significant differences in their Mandarin use, because their Mandarin input will have been very similar. However, when we found that there were some collocations used by one or two groups significantly more often than the other groups, our hypotheses are challenged. Even if the difference is a result of a different preference/tendency in the participants' discourse use, not in their lexico-grammatical usage, it is still odd that a particular preference/tendency occurs significantly more often or less often in a particular group of participants with a similar linguistic background in comparison with groups with different backgrounds which theoretically ought not to influence their use of Mandarin. Although the main core of the current study is about the differences in the lexico-grammatical use and whether the lexico-grammatical differences can be explained by the influence of the different language input to which these participants had been exposed, what we have found in our investigations above with regard to differences in discourse preference suggests that there, somehow, might be a cross-language influence on how these participants use English and Mandarin with respect to not only the lexico-grammatical aspects of the languages but also the discourse aspects of the languages (e.g the use of *I just*/我就 [I just] and *I think*/我觉得 [I think]). It would be a worthwhile study to explore these differences of discourse preference in more detail, but it is impossible to cover this part in this thesis. We will focus on certain collocations which may potentially have different lexico-grammatical use, such as the combination of the noun 时候 [time] with the pronoun 你 [you] and the combination of the verb 让 [let] with the pronoun 你 [you]. We will also discuss the differences of discourse preferences between groups as they arise.

时候(shíhou) 你(nǐ) [time/when you]

In our investigation on the collocation 时候 你 [time you] and the word 时候 [time], it is found that not only was the collocation 时候 你 [time you] used more frequently by the LIVT and TESOL participants, but the word 时候 [time] was also used more frequently by the LIVT and TESOL participants than by the NTESOL participants. (See Table 5.18.) Despite of the different frequencies, when we checked the collocates of the word 时候 [time] and the collocation 时候 你 [time you], it is found that their collocates did not differ greatly. (See Table 5.19.)

时候(shíhòu) 你(nǐ)	LLR	Sig.		
LIVT - TE	1.513	0.219		+
LIVT - NTE	12.649	**0.000**	***	+
TE - NTE	5.080	**0.024**	*	+
时候(shíhòu)	LLR	Sig.		
LIVT - TE	1.37	0.242		+
LIVT - NTE	17.39	**0.000**	***	+
TE - NTE	8.32	**0.004**	**	+

Table 5.18 Log-likelihood Ratios: 时候(shíhòu) and 时候(shíhòu) 你(nǐ) in LIVT, TESOL and NTESOL

We found that the word 时候 [time] frequently appeared in the collocations 有/有的 时候 [similar to *sometime* in English] and 那 时候 [that time]. However, the most frequent collocation was 的 时候 [of time/when]. The collocation 的 时候 [of time/when] functions in much the same way as the word *when* in English. Like the *when*-clause, the collocation 的 时候 is used to indicate a specific period of time or condition, as shown in the sample below:

写 作 文 的 时候 = when (a person) writes an essay
xiě zuòwén de shíhou
write essay de(of) time/period

	LIVT			TESOL			NTESOL		
	Freq.	**Per 1000 words**		**Freq.**	**Per 1000 words**		**Freq.**	**Per 1000 words**	
时候(shíhou)	246	6.38		202	5.72		179	4.25	
时候 你(shíhou nǐ)	27	0.70		17	0.48		8	0.19	
R1 Collocates of 时候(shíhou) (TOP 5)	R1	Freq	Per 1000 words	R1	Freq	Per 1000 words	R1	Freq	Per 1000 words
	我 (wǒ)	31	0.80	你 (nǐ)	17	0.48	我 (wǒ)	20	0.47
	你 (nǐ)	27	0.70	我 (wǒ)	13	0.37	就 (jiù)	16	0.38
	就 (jiù)	15	0.39	会 (huì)	12	0.34	会 (huì)	11	0.26
	可能 (kě néng)	12	0.31	就 (jiù)	12	0.34	是 (shì)	9	0.21
	是 (shì)	11	0.29	就是 (jiù shì)	11	0.31	你 (nǐ)	8	0.19
Patterns of 时候(shíhou) (TOP 5)	N	L1	Centre	R1					
	1	的	时候	我					
		de	shíhou	wǒ					
		of	time/period	I					
	2	那		你					
		nà		nǐ					
		that		you					
	3	个		就					
		gè		jiù					
		classifier		just					
	4	有的		可能					
		yǒude		kěnéng					
		(sometime)		maybe					
	5	有		是					
		yǒu		shì					
		(sometime)		to be					
	N	L1	Centre	R1					
	1	的	时候	你					
		de	shíhou	nǐ					
		of	time/period	you					
	2	那		我					
		nà		wǒ					
		that		I					
	3	些		会					
		xiē		huì					
		classifier		can					
	4	个		就					
		gè		jiù					
		classifier		just					
	5	有的		就是					
		yǒude		jiùshì					
		(sometime)		just					
	N	L1	Centre	R1					
	1	的	时候	我					
		de	shíhou	wǒ					
		of	time/period	I					
	2	那		就					
		nà		jiù					
		that		just					
	3	个		会					
		gè		huì					
		classifier		can					
	4	有的		是					
		yǒude		shì					
		(sometime)		to be					
	5	什么		你					
		shénme		nǐ					
		what		you					
Patterns of 时候你(shíhou nǐ) (TOP 5)	N	L1	Centre	R1					
	1	的	时候你	可能					
		de	shíhounǐ	kěnéng					
		of	time you	maybe					
	2			就是					
				jiùshì					
				just					
	3			可以					
				kéyǐ					
				can					
	N	L1	Centre	R1					
	1	的	时候你	可以					
		de	shíhounǐ	kéyǐ					
		of	time you	can					
	N	L1	Centre	R1					
	1	的	时候你						
		de	shíhounǐ						
		of	time you						

Table 5.19 Use of 时候(shíhou) 你(nǐ) and 你(nǐ) in LIVT, TESOL and NTESOL

With respect to the use of the collocation 时候 你 (shíhou nǐ) [time you], it is mostly used in the pattern 的 时候 你 会/可能/可以 (de shíhou nǐ huì kěnéng kěyǐ) [~of the period/when~ you can/may], as in the instances shown below in Table 5.20. The heavy use of the word 时候 (shíhou) [time] and the collocation 时候你 (shíhou nǐ) [time you] suggests that the LIVT and TESOL participants might tend to give description of the time or condition more

frequently than the NTESOL participants, especially when the subject was the second person 你 [you]. The use by the three groups with regard to the word 时候 [time] and the collocation 时候 你 [time you], however, did not show any obvious differences.

LIVT	或 许 是 说 写 作 的 时 候 你 可能 运 , 会 运 用 到
	huòxǔ shì shuō xiězuò de shíhou nǐ kěnéng yùn , huì yùnyòng dào
	It is said that maybe when you write an essay, you can use [it].
TESOL	你 在 写 的 时 候 你 可以 引经据典 嘛
	nǐ zài xiě de shíhou nǐ kéyǐ yǐnjīngjùdiǎn ma
	When you are writing, you can quote from the classics.
NTESOL	写 笔记 的 时 候 你 会 用 简体字 带 过
	xiě bǐjì de shíhou nǐ huì yòng jiǎntǐzì dàiguò
	When you take a note, you will use simplified characters.

Table 5.20 Instances of the use of 时候(shíhou) 你(nǐ) in LIVT, TESOL and NTESOL

Although we did not find that participants were prone to use the word 时候 [time] and the collocation 时候 你 [time you] noticeably differently in terms of their lexico-grammatical usage, we did find that the participants appeared to behave differently in terms of their discourse preferences. We assume that the significant high use of 时候 你 [time you] in the LIVT and TESOL groups indicates that the LIVT and TESOL participants tended to provide description of the time or condition more frequently than the NTESOL participants, especially when the subject was the second person 你 [you]. Now we will look more closely at the collocation 的 时候 [when] with 你 [you] with a view to test whether our assumption is correct.

In Table 5.21, we can see that the LIVT and TESOL participants used the words 的 时候 [when] with a similar frequency (3.60 times per 1000 words in the LIVT group and 3.57 times per 1000 words in the TESOL group.) Both of these groups used the words 的 时候 [when] slightly more often than the NTESOL group (2.8 times per 1000 words.) When the use of 的 时候 [when] involved the pronoun 你 [you], it is found that the LIVT participants used 的 时候 你 [when you] more often than the TESOL and NTESOL participants, especially much more often than the NTESOL participants. The use of 的 时候 你 [when you] occurred 0.31 times per 1000 words in the LIVT group, 0.20 times per 1000 words in the TESOL group and 0.12 times per 1000 words in the NTEOSL group. The LIVT participants used 的 时候 你 [when you] twice as often as the NTESOL

participants. Thus we can see that the claim, in which we stated the LIVT and TESOL participants tended to provide description of the time or condition more frequently than the NTESOL participants, especially when the subject is the second person 你 [you], is correct.

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
的 时候(de shíhou) [when]	139	**3.60**	126	**3.57**	118	**2.80**
的 时候(de shíhou)你(nǐ) [when you]	12	**0.31**	7	**0.20**	5	**0.12**
when	178	6.17	212	7.14	134	5.75
when you	29	**1**	16	**0.54**	9	**0.39**

Table 5.21 Use of 的 时候(de shíhou) [when] and 的 时候(de shíhou)你(nǐ) [when you] in LIVT, TESOL and NTESOL

In the previous chapter, we also found that the LIVT participants tended to use the collocation *when you* significantly more often than the other participants. We took both the English use of *when* and *when you* into comparison, and we found a similar discourse preference in the Mandarin and English use of these expressions. The LIVT and TESOL participants also tended to use the word *when* for indicating the time/condition slightly more often than the NTESOL participants. When the word *when* was used with the second person pronoun *you*, the LIVT participants used the collocation *when you* more often than the TESOL and NTESOL participants. The LIVT participants also used *when you* over twice as often as the NTESOL participants. What we have found here is that, similarly to what we found in the use of *I just*/我 就 [I just] and *I think*/我 觉 得[I think], the LIVT participants showed similar discourse preferences in both their English and Mandarin. However, in the case of the TESOL and NTESOL participants, there was no such similarity between their English and Mandarin uses. This suggests that our null hypotheses (hypothesis 3 and hypothesis 4) are greatly challenged.

These differences, although they are discourse differences, may still indicate that different language input potentially affect how participants used their language at the discourse level. We have found that the LIVT, TESOL and NTESOL groups had learnt how to use certain English words and phrases which serve similar discourse uses to those served by Mandarin words. However, the TESOL and NTESOL participants had not been primed as strongly as the LIVT participants had. This finding is potentially compatible with lexical priming theory although the observation was made across two languages, and there was not too much discussion about cross language effects in current existing studies of lexical priming. If a person has his/her own discourse preference for emphasising or avoiding emphasis in his/her talk, this discourse preference should be consistent no matter which language this person uses. But what we have seen here is that for the TESOL and NTESOL

participants who had had fewer opportunities to encounter and use English than the LIVT participants had had, the strength of their discourse preference was somehow weaker when they used English. This finding demonstrates that the degree of people's exposure to the language input (English) appears to influence the strength of people's priming of word/word sequences, and what is influenced is not simply the priming of lexico-grammatical use (collocation/colligation and semantic association) but also the strength of discourse use.

让 你(rang nǐ) [let you]

The collocation ràng 让 n ǐ 你 [let you] was used by the LIVT participants more frequently than the TESOL and NTESOL participants, whereas the overall frequencies of the word ràng 让 [let] on its own did not differ much across the three groups. (See Table 5.22.)

让你(ràng nǐ)	LLR	Sig.		
LIVT - TE	5.844	**0.016**	*	+
LIVT - NTE	5.214	**0.022**	*	+
TE - NTE	0.066	0.798		-
让(ràng)	LLR	Sig.		
LIVT - TE	0.67	0.413		+
LIVT - NTE	0.32	0.574		+
TE - NTE	0.08	0.774		-

Table 5.22 Log-likelihood Ratios: 让(ràng) and 让你(ràng nǐ) in LIVT, TESOL and NTESOL

The word ràng 让 can be used as a verb meaning 'to let/have someone do something' (See sample [A] in Table 5.23) or as a preposition suggesting something is done by someone (See sample [B] in Table 5.23).

[A]	我 让 他 去 打 扫 客 厅 。
	wǒ ràng tā qù dásǎo kètīng。
	I let him clean the living room.
[B]	这 孩 子 让 你 宠 坏 了。
	zhè háizi ràng nǐ chǒnghuài le。
	This kid was spoiled by you.

Table 5.23 Examples of the use of 让(ràng)

Both of the use indicate that there is a relationship of influence between two agents. Therefore, it is not surprising to see that the word 让(ràng) [let] is followed by pronouns and nouns, as shown in Table 5.24. We found that all the participants used similar R1 collocates of 让(ràng) [let], and the pronoun 你(nǐ) [you] was common to all the groups. However, there is one thing worth noting when we observed the patterns of use of 让你(ràng nǐ) [let you] in the three groups. In the instances of 让(ràng) 你(nǐ) [let you], it is found that the LIVT participants used 让(ràng) 你(nǐ) [let you] in the collocation 让(ràng) 你(nǐ) 去(qù) [let you 'to go'] more often than the TESOL and NTESOL participants. (See the patterns of 让你(ràng nǐ) [let you] in Table 5.24.)

	LIVT			**TE**			**NTE**		
	Freq.	**Per 1000 words**		**Freq.**	**Per 1000 words**		**Freq.**	**Per 1000 words**	
让(ràng)	80	2.07		64	1.81		80	1.90	
让你(ràng nǐ)	24	0.62		9	0.25		13	0.31	
R1 Collocates of 让(ràng) (TOP 5)	R1	Freq	Per 1000 words	R1	Freq	Per 1000 words	R1	Freq	Per 1000 words
	你(nǐ)	24	0.62	我(wǒ)	12	0.34	他们(tā men)	18	0.43
	我(wǒ)	13	0.34	你(nǐ)	9	0.25	我(wǒ)	14	0.33
	他(tā)	8	0.21	他们(tā men)	8	0.23	你(nǐ)	13	0.31
	人家(rén jiā)	5	0.13	人家(rén jiā)	8	0.23	他(tā)	9	0.21
	学生(xué shēng)	4	0.10	学生(xué shēng)	3	0.08	人家(rén jiā)	6	0.14
Patterns of 让你(ràng nǐ) (TOP 5)	N	L1	Centre	R1	N	L1	Centre	R1	N	L1	Centre	R1
	1	然后	让 你	去	1	能 够	让 你		1	就 是	让 你	
		ránhòu	ràngnǐ	qù		nénggòu	ràngnǐ			jiùshì	ràngnǐ	
		then	let you	go		can	let you			just	let you	
	2	时间		的	2	可以			2	可以		
		shíjiān		de		kéyǐ				kéyǐ		
		time		of		can				can		

Table 5.24 Use of 让你(ràng nǐ) and 你(ràng) in LIVT, TESOL and NTESOL

The word 去(qù) can be a verb means 'to go to' (See Sample [C] in Table 5.25.) But it is also used as a grammatical particle indicating something is in process. (See Sample [D].) The most common use of the word 去(qù) is as a supplementary verb before another verb. (See Sample [E].) It adds an extra tone of activeness to the followed verb, and the meaning of the sentence will not differ greatly without it. (See Sample [F].) In the instances of the collocation 让(ràng) 你(nǐ) [let you] in the LIVT group, 7 out of 24 instances are found in the pattern 让(ràng) 你(nǐ) 去(qù) + verb [let you to go to + verb]. There

is no such an instance found in the TESOL group, and only two instances found in the NTESOL group. That is to say, the use of 让(ràng) 你(nǐ) 去(qù) + verb [let you to go to + verb] occurrs 0.18 times per 1000 words in the LIVT group, 0 time per 1000 words in the TESOL group and 0.05 times per 1000 words in the NTESOL group. The LIVT participants used the pattern 让(ràng) 你(nǐ) 去(qù) + verb [let you to go to + verb] more often than the TESOL and NTESOL participants. The difference was marked. Therefore, we need to investigate further the use of the patterns 让(ràng) * 去(qù) + verb [let * to go to + verb], 让(ràng) * + verb [let * + verb] and 去(qù) + verb [to go to + verb]. (See Table 5.26.)

[C]	我 去 学 校 。
	wǒ qù xuéxiào。
	I go to school.
[D]	他 上 班 去 了。
	Tā shàngbān qù le。
	He went to work (and is working now).
[E]	你 可以 去 表 达 想 法。
	Nǐ kéyǐ qù biǎodá xiángfǎ。
	You can (go and) express your thought.
[F]	你 可以 表 达 想 法。
	Nǐ kéyǐ biǎodá xiángfǎ。
	You can express your thought.

Table 5.25 Examples of the use of 去(qù) [to go (to)]

	LIVT			**TE**			**NTE**		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
让(ràng) 你(nǐ)	24	100	0.62	9	100	0.25	13	100	0.31
让(ràng) 你(nǐ) 去(qù) + V	7	29.17	**0.18**	0	0.00	**0.00**	2	15.38	**0.05**
让(ràng) 你(nǐ)+ V	6	25.00	0.16	4	44.44	0.11	4	30.77	0. 09
让(ràng)	80	100	2.07	64	100	1.81	80	100	1.90
让(ràng) * 去(qù) + V	9	**11.25**	**0.23**	3	**4.69**	**0.08**	4	**5.00**	**0.09**
让(ràng) * + V	32	**40.00**	**0.83**	26	**40.63**	**0.74**	48	**60.00**	**1.14**
去(qù)	229	100	5.94	244	100	6.91	234	100	5.56
去(qù) + V	183	**79.91**	**4.74**	178	**72.95**	**5.04**	183	**78.21**	**4.34**

Table 5.26 Use of 让你(ràng nǐ), 让(ràng) and 去(qù) in LIVT, TESOL and NTESOL

In Table 5.26, it is found that in the use of 去(qù) + verb [to go (to) + verb] there were no great differences between the groups in terms of the frequencies of this pattern. The LIVT participants used 去(qù) + verb [to go to + verb] 4.74 times per 1000 words. The TESOL participants used 去(qù) + verb [to go to + verb] 5.04 times per 1000 words, and the NTESOL participants used [to go to + verb] 4.34 times per 1000 words. It is also found that the use of the word 去(qù) [to go to] followed by the verbs is the primary use of the word 去(qù) [to go to] in all the groups. This finding suggests that these participants did not differ greatly in terms of the use of the word 去(qù) [to go to]. However, when investigating the use of 让(ràng) [let] and 去(qù) [to go to], the LIVT participants behaved differently from the TESOL and NTESOL participants. When the LIVT participants used the word 让(ràng)[let], they were more prone to use the word 去(qù) [to go to] in front of the verbs. The pattern 让(ràng) * 去(qù) + verb [let * to go to] occurred 0.23 time per 1000 words (11.25% of the total use of 让(ràng)[let]) in the LIVT group, 0.08 times per 1000 words (4.69% of the total use of 让(ràng)[let]) in the TESOL group and 0.09 times per 1000 words (5% of the total use of 让(ràng)[let]) in the NTESOL group. The LIVT participants used the pattern 让(ràng) * 去(qù) + verb [let * to go to + verb] twice as often as the TESOL and NTESOL participants. This finding suggests that the LIVT participants, although their use for the words 让(ràng) [let] and 去(qù) [to go to] did not differ greatly from the other groups in terms of their lexico-grammatical features, did have a tendency to add an mark of activeness with the word 去(qù) [to go to] when they used the word 让(ràng) [let]. However, when we take the parallel use in English - the use of the patterns *let * + verb* and *let * to + verb* into our investigation - the evidences of another cross-language effect on people's use of language in their lexico-grammatical use are found.

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
让(ràng) * + V	32	0.83	26	0.74	48	1.14
let * + V	6	0.21	5	0.17	9	0.39
让(ràng) * 去(qù) + V	9	**0.23**	3	**0.08**	4	**0.09**
*let * to + V	0	**0.00**	3	**0.10**	1	**0.04**

Table 5.27 Use of 让(ràng)* +*verb*, 让(ràng)*去(qù) +*verb*, *let* *+ *verb* and *let* * *to* +*verb* in LIVT, TESOL and NTESOL

In Table 5.27, we find that the LIVT, TESOL and NTESOL participants used the pattern *let* * + *verb* in English in a similar way to the use of the pattern 让 * 去 + verb [let * to go to + verb] in Mandarin. The LIVT and TESOL participants used the Mandarin pattern 让 * 去 + verb [let * to go to + verb] less often than the NTESOL participants, and the LIVT participants used this pattern slightly more often than the TESOL participants. The use of the English pattern *let* * + *verb* was the same. The LIVT and TESOL participants used the English pattern *let* * + *verb* less often than the NTESOL participants, and the LIVT participants used this pattern slightly more often than the TESOL participants.

However, in respect of the use of the Mandarin pattern 让 * 去 + verb [let * to go to + verb], it is found that the differences between the three groups became distinct. The LIVT participants used the Mandarin pattern 让 * 去 + verb [let * to go to + verb] more often than the other participants, which we interpreted as a means of adding an extra emphasis to the action of the agent. This kind of usage is comparatively rarer in the TESOL and NTESOL groups. However, when we check whether there is any use of the English pattern *let * to + verb*, which is an uncommon use in English, in the three groups, it is found that the TESOL and NTESOL participants had this kind of use. There are 3 occurrences of the use *let * to + verb* in the TESOL data, and 1 occurrence of the use of *let * to + verb* in the NTESOL data. Such an uncommon English pattern is not found in the LIVT data.

The possible explanations why some participants in the TESOL and NTESOL groups (although not many) used the rare English pattern *let * to + verb* are twofold. First, these participants might have been primed in English strongly to associate two verbs with the word *to* (*verb + infinitive-to + verb*), and they might have been primed less strongly in their English for expressions where there was no *to* between two verbs. Therefore, when they used the verb *let* and another verb, they might have been primed to use the pattern *let * to + verb*. The LIVT participants, instead, might have been primed to use the pattern *let* * + verb strongly, so they did not use the pattern *let * to + verb* in their English. The second possible explanation might be that the TESOL and NTESOL participants were influenced by the use of 让 * 去 + verb [let * to go to +verb] in Mandarin. The word 去 [to go to], as mentioned early, is commonly used as a supplementary verb before another verb in Mandarin. The word 去 [to go to] contains a sense of direction similar to the English word *to*. It might be that the participants who used the pattern *let * to + verb* used a similar priming of 去 [to go to] in Mandarin. That is to say, their English priming of the lexico-grammatical use in this use was not strong enough, so their Mandarin priming was used in its place. In the case of the LIVT

participants, however, their English priming for the lexico-grammatical use in this use was strong enough, and therefore, although they used the pattern 让 * 去 + verb [let * to go to +verb] often in their Mandarin, their English priming was not affected strongly. Either way, what we have found proves that our null hypotheses in Mandarin are incorrect, in that there are noticeable differences between the LIVT group and the TESOL and NTESOL groups. The finding also proves that the LIVT participants were primed differently in the use of English from the TESOL and NTSOL participants and vice versa. Furthermore, different degrees of exposure to English input are the likely explanation for this difference.

5.3.2 R1 Collocates of 你(nǐ) [you]

In this section, we move to investigate the R1 collocates of 你[you]. We can see from Table 5.28 that the three groups shared many of the same R1 collocates in their top 10 most frequent R1 collocates lists. The difference occurred in the words 你 [you], 有[have], 写 [to write], 不[no/not] and 是[to be]. The LIVT group has 你[you] as the R1 collocate of 你[you] frequently and it is not seen in the other two groups' lists. The TESOL group has 有 [to have/to exist] as a R1 collocate and it is not seen in the other two groups' lists. The NTESOL group has the words 是 [to be] and 写 [to write] as frequent R1 collocates and it is also not seen in the other groups' lists. The LIVT and TESOL groups have the word 不[no/not] as the frequent R1 collocates, and it is not seen in the NTESOL group's list. In order to see whether there are any significant differences in terms of the frequencies of these R1 collocates, these collocations are examined with log-likelihood ratios and presented in Table 5.29.

	LIVT	1014	26.29	TESOL	632	17.90	NTESOL	637	15.12
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	要	67	1.74	要	75	2.12	的	43	1.02
2	可能	56	1.45	的	53	1.50	就	37	0.88
3	的	50	1.30	就	37	1.05	要	35	0.83
4	就	45	1.17	说	33	0.93	可以	23	0.55
5	会	37	0.96	可以	32	0.91	可能	21	0.50
6	说	33	0.86	会	24	0.68	在	20	0.47
7	可以	33	0.86	在	20	0.57	会	19	0.45
8	你	31	0.80	可能	14	0.40	写	17	0.40
9	不	27	0.70	有	14	0.40	说	16	0.38
10	在	23	0.60	不	14	0.40	是	16	0.38

Table 5.28 R1 Collocates of 你(nǐ) in LIVT, TESOL and NTESOL

LIVT - NTE						LIVT - TE						TE - NTE					
Over						Over						Over					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
你	可能	19.768	0.000	***	+	你	可能	23.496	0.000	***	+	**你**	**要**	**22.723**	**0.000**	***	+
你	**要**	**13.254**	**0.000**	***	+	你	你	9.551	0.002	**	+	**你**	**说**	**9.371**	**0.002**	**	+
你	你	9.270	0.002	**	+												
你	**说**	**7.629**	**0.006**	**	+												
你	会	7.599	0.006	**	+												
你	不	3.898	0.048	*	+												

Table 5.29 Log-likelihood Ratios for R1 collocates of 你(nǐ) in LIVT, TESOL and NTESOL

In Table 5.29, it is found that the LIVT participants used the collocations 你 可能 [you might/maybe] and 你 你 [you you] more often than the TESOL and NTESOL participants. The LIVT participants also used the collocations 你 会 [you can] and 你 不[you no/not] more often than the NTESOL participants. The LIVT and TESOL participants used the collocation 你 要 [you want] and 你 说 [you say] significantly more often than the NTESOL participants. We will not talk about the use of repetition (你 你 [you you]) and negation (你 不[you no/not]) in this study. The words 可能 [might/maybe] and 会 [can] are often used as modal verbs/adverbs to express modality (possibility), but although it would be a valuable topic to see how the three groups used modal verbs/adverbs to express modality in their speech, we will not be covering this area in the current study. In the following section, we will take a close look at the use of the verbs 要 [want] and 说 [say] with the pronoun 你 [you] in order to see whether there are more findings of noticeable differences beside that of the different frequencies in the three groups.

你(nǐ) 要(yào) [you want/going to/must]

Inspecting the use of the collocation 你 要 [you want/going to/must], it is found that the LIVT and TESOL participants used the collocation 你要 [you want/going to/must] significantly more frequently in their speech than did the NTESOL participants. There are 1.76 times per 1000 words of the use of 你要 [you want/going to/must] in the LIVT data, 2.12 times per 1000 words of the use of 你要 [you want/going to/must] in the TESOL data and 0.85 times per 1000 words of the use of 你要[you want/going to/must] in the NTESOL data. The LIVT and TESOL participants tended to use the collocation three time/twice more often than the NTESOL participants. (See Table 5.30.)

你要 (nǐ yào)	LLR	Sig.		
LIVT - TE	1.403	0.236		-
LIVT - NTE	13.254	**0.000**	***	+
TE - NTE	22.723	**0.000**	***	+

Table 5.30 Log-likelihood Ratios: 你要(nǐ yào) in LIVT, TESOL and NTESOL

When examining the instances of the use of 你 要 [you want/are going to/must] in the groups, it is found that the collocation 你要 [you want/are going to/must] is commonly accompanied by the adverb (就是) and the connectives (然后 [and then], 因为 [because] and 如果 [if]) in its L1 position. It is also commonly accompanied by the adverb 怎么/怎么样 [how] in its R1 position. (See 5.31.) The pattern 然后 你 要 怎么 [and then, you want/are going to, how] (and then how you are going to/want to)] is a common pattern shared by all the three groups. There are no distinct differences in the collocations/lexico-grammatical use of 你要 [you want/are going to/must] associated with the patterns. However, as we mentioned previously, the word 要 has three different meanings in Mandarin. The word 要 is often used as a verb meaning 'want/request'. It can also be a

modal verb meaning ‘must/should’ or a future auxiliary meaning ‘will/*BE* going to’. We will now examine whether there are any noticeable differences in the use of these three kinds of 要(yào) .

	LIVT			TE			NTE		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
要 (yào)	298	100	7.73	375	100	10.62	333	100	7.91
你要 (nǐ yào)	68	22.82	**1.76**	75	20.00	**2.12**	36	10.81	**0.85**
Patterns of 你要 (nǐ yào) (TOP 5)	N / L1 / Centre / R1 1 / 然后 / 你要 / 怎麼 ránhòu / nǐyào / zěnme then / you want / how 2 / 就是 / / 怎麼样 jiùshì / / zěnmeyàng just / / how 3 / 因为 / / yīnwèi / / because / / 4 / 如果 / / rúguǒ / / if / / 5 / 表达 / / biǎodá / / express / /			N / L1 / Centre / R1 1 / 时候 / 你要 / 怎麼 shíhou / nǐyào / zěnme time/period / you want / how 2 / 就是 / / 自己 jiùshì / / zìjǐ just / / oneself 3 / 然后 / / 怎麼样 ránhòu / / zěnmeyàng then / / how 4 / 如果 / / 引用 rúguǒ / / yǐnyòng if / / to quote 5 / 所以 / / suóyǐ / / so / /			N / L1 / Centre / R1 1 / 然后 / 你要 / 怎麼 ránhòu / nǐyào / zěnme then / you want / how 2 / 就是 / / jiùshì / / just / / 3 / 如果 / / rúguǒ / /		

Table 5.31 Use of 要(yào) and 你要(nǐ yào) in LIVT, TESOL and NTESOL

	LIVT			TE			NTE		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
你 要 (nǐ yào)	68	100	1.76	75	100	2.12	36	100	0.85
你 要 (nǐ yào) (to want/to request)	23	**33.82**	0.60	30	**40.00**	0.85	22	**61.11**	0.52
你 要 (nǐ yào) (must/should)	42	**61.76**	1.09	45	**60.00**	1.27	14	**38.89**	0.33
你 要 (nǐ yào) (will/going to)	3	4.41	0.08	0	0.00	0.00	0	0.00	0.00

Table 5.32 Use of 要 (yào) in 你要 (nǐ yào) in LIVT, TESOL and NTESOL

In Table 5.32, we can see that, in its total use of the collocation 你要(nǐ yào), the LIVT participants used the word 要(yào) in the meaning of ‘want/request’ 23 times (33% of the total use of 你要(nǐ yào)), the word 要(yào) in the meaning of ‘must/should’ 42 times (61.76% of the total use of 你要(nǐ yào),) and the word 要(yào) in

the meaning of ‘will/*BE* going to’ 3 times (4.41 % of the total use of 你要). The TESOL participants used the word 要 in the meaning of ‘want/request’ 30 times (40% of the total use of 你要,), the word 要 in the meaning of ‘must/should’ 45 times (60% of the total use of 你要,). There was no use of the word 要 with the meaning of ‘will/*BE* going to’ in the instances of 你要 found in the TESOL group. The NTESOL participants used the word 要 in the meaning of ‘want/request’ 22 times (61.11% of the total use of 你要), the word 要 in the meaning of ‘must/should’ 14 times (38.89% of the total use of 你要) There was again no use of the word 要 with the meaning of ‘will/going to’ in the NTESOL group in the context of their use of 你要. These findings suggest that the participants from all the groups used the collocation 你要 in a similar way as regards its lexico-grammatical use (its collocations/colligations). However, the LIVT and TESOL participants tended to have more discussion on what a person should do, and the NTESOL group tended to talk about what a person would want to do. (Here the pronoun 你 was used as a general pronoun.) The finding that the LIVT and TESOL participants used the collocation 你要 [you want/*BE* going to/must] significantly more often than the NTESOL participants has challenged our null hypotheses. Although we did not find great differences in the collocations/colligations of 你要, we did find that the LIVT and TESOL participants and the NTESOL participants used the combination 要(你要) in different ways in the discourse. It is difficult to say the differences result from their having different English input. This difference is probably caused by the different perspective that these groups of participants took on the educational topics. What EFL learners with different language learning backgrounds think of the school education (in learning English/Mandarin) would be an interesting area to explore, but it would take us away from our research questions.

你(nǐ) 说(shuō) [you say]

Looking at the collocation 你 说 [you say], it is found that the LIVT and TESOL participants used the collocation 你 说[you say] significantly more often than did the NTESOL participants. (See Table 5.33.)

你说 (nǐ shuō)	LLR	Sig.		
LIVT - TE	0.122	0.726		-
LIVT - NTE	7.629	**0.006**	**	+
TE - NTE	9.371	**0.002**	**	+

Table 5.33 Log-likelihood Ratios: 你说 (nǐ shuō) in LIVT, TESOL and NTESOL

The LIVT participants used the collocation 你 说 [you say] 0.87 times per 1000 words. The TESOL participants used the collocation 你 说 [you say] 0.95 times per 1000 words. The NTESOL participants used the collocation 你 说 [you say] 0.39 times per 1000 words. The collocation 你 说 [you say] was used more than twice as often by the LIVT and TESOL participants as by the NTESOL participants. (See Table 5.34.)

	LIVT		**TE**		**NTE**	
	Freq.	**Per 1000 words**	**Freq.**	**Per 1000 words**	**Freq.**	**Per 1000 words**
你说 (nǐ shuō)	33	**0.87**	33	**0.95**	16	**0.39**
L1 Collocates of *你说 (nǐ shuō)*	**告诉 [tell] 3** **然后 [then] 2** 嗯 [eh] 2 喔 [oh] 2 啊 [ah] 2		**跟 [to] 3** 嗯 [eh] 3 看法 [viewpoint] 2 呃 [uh] 2		喔 [oh] 3 **跟 [to] 2** 嗯 [eh] 2	
R1 Collocates of *你说 (nǐ shuō)*	你 [you] 4 我 [I] 3 中文 [Chinese] 2		**的 [of/DE] 3** 我 [I] 2 就是 [just] 2		我 [I] 3 诗词 [poetry] 2	
Patterns of *你说 (nǐ shuō)* (TOP 5)	N / L1 / Centre / R1 1 / 告诉 / 你 说 / 你 / gàosù / nǐ shuō / nǐ / tell / you say / you 2 / 然后 / / 我 / ránhòu / / wǒ / then / / I 3 / / / 中文 / / / zhōngwén / / / Chinese		N / L1 / Centre / R1 1 / 跟 / 你 说 / 的 / gēn / nǐ shuō / de / to / you say / of 2 / 看法 / / 我 / kànfǎ / / wǒ / viewpoint / / I 3 / / / 就是 / / / jiùshì / / / just		N / L1 / Centre / R1 1 / 跟 / 你 说 / 我 / gēn / nǐ shuō / wǒ / to / you say / I 2 / / / 诗词 / / / shīcí / / / poetry	

Table 5.34 Use of 你说 (nǐ shuō) in LIVT, TESOL and NTESOL

In our investigation on the use of 你 说 [you say], some differences between groups are found. The collocates of 你 说 [you say] are different. In the use of L1 collocates of 你 说 [you say], the LIVT participants used the collocations 告诉你说 [tell you say] and 然后你说 [and then you say] more often than the other participants. The TESOL and NTESOL participants had the collocation 跟你说 [to you say (say to you)] that the LIVT participants did not use very often. There is only one instance of 跟你说 [to you say (say to you)] found in the LIVT group, and likewise only one instance of 告诉你说 [tell you say] found in the NTESOL group. There is no use of 然后你说 [and then you say] found in the TESOL and NTESOL groups. In the use of R1 collocates of 你说 , it found that the TESOL participants used the collocation 你说 的 [you say of (DE)] more often than the other participants. There is only one instance of 你说 的 [you say of(DE)] found in the LIVT group and none in the NTESOL group. From these instances, it can be seen that the use of 你说 的 [you say of (DE)] is similar to the meaning of 'what you said/mentioned of ~ is' and is usually followed by the word 是 [to be]. We will discuss this use in chapter 6. The use of different collocates and patterns of 你说[you say] across the three groups suggests that the groups might differ in their lexico-grammatical use of 你说 [you say]. Therefore, we will take a closer look at how our participants used 你说 [you say] in detail.

From the observation of all the instances of 你说 [you say] in the three groups, it is found that there are four types of use. The first type is to use 你说 [you say] in an interrogative sentence. The participants used 你说 [you say] to check whether they understood what the speaker (interviewer) said correctly. (See Example 1 in Table 5.35.) The second type of use was similar to the first one, but in declarative form. The participants used 你说 [you say] to reference what the speaker (interviewer) just said, so that they could respond to the speaker. (See Example 2 in Table 5.35.) The first and the second types of use had connectives or adverbs as their L1 colligates. The third type of use involved the use of the preposition 跟 [to] in the L1 position. The collocation 跟 你 说 means 'tell you'. (See Example 3.) The fourth type involved the use of a verb as a R1 colligate with the pronoun 你

[you], and the word shuō 说[say] is used as a emphasiser. The word shuō 说 can be omitted without influencing the meaning of the sentence. (See Example 4.)

1	# **你 说** 现 在 我 的 状 况 **吗** ?
	# Nǐ shuō xiànzài wǒ de zhuàngkuàng ma ?
	# **Is it** 'my current situation' **that you are saying**?
2	# 我 觉得 不 太 妥 , 而且 **你 说** 效率 , 效率 跟 写作 , 欸 , 写作 的 话 …
	# Wǒ juédé bú tài tuǒ , érqiě nǐ shuō xiàolǜ , xiàolǜ gēn xiězuò , èi , xiězuò de huà …
	# I think it's not appropriate. And **you said** efficiency and writing, eh, if it is writing …
3	就 像 我 刚 刚 **跟 你 说** , 就是 蔡 康 永 啊。
	Jiù xiàng wǒ gānggāng gēn nǐ shuō , jiùshì càikāngyǒng ā 。
	Just as what I **told you**, it's 蔡康永 (name).
4	他 不 会 呃 , **要 求 你 说** 要 写 出 什 么 样 的 东西。
	Tā bú huì è , yāoqiú nǐ shuō yào xiěchū shénme yàng de dōngxi
	He's not going to **ask you (saying that)** (you) must write something down.

Table 5.35 Examples of the four types of the use of 你说 (nǐ shuō) in LIVT, TESOL and NTESOL

		LIVT			TE			NTE		
		Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
	nǐ shuō 你说 [you say]	33	100	0.86	33	100	0.93	16	100	0.38
1	**你说 NP/[S]** ma 吗? (Do you mean ~?)	**19**	**57.58**	0.49	**20**	**60.61**	**0.57**	**10**	**62.50**	0.24
2	**你说 NP/[S]** (what you said is~)	**7**	**21.21**	0.18	**10**	**30.30**	**0.28**	**3**	**18.75**	0.07
3	Prep.+ 你说**NP/[S]** (talk to you/tell you)	1	3.03	0.03	3	9.09	0.08	2	12.50	0.05
4	Verb + 你说**NP/[S]** (shuō 说 as an emphasiser 'saying')	**6**	**18.18**	0.16	**0**	**0.00**	**0.00**	**1**	**6.25**	0.02

Table 5.36 The four types of the use of 你说 (nǐ shuō) used by LIVT, TESOL and NTESOL

In the total instances of the collocation nǐ shuō 你说 [you say], the first and the second types of use were frequently used by all the participants. Over half of the instances of nǐ shuō 你说 [you say] were the

instances of the first type of use in all the groups. (LIVT: 57.58%; TESOL: 60.61%; NTESOL: 62.50%) The second type of use accounts for around 20% of the total use of 你说 [you say] in the LIVT and NTESOL groups. It accounts for 30.30% of the total use of 你说 [you say] in the TESOL group. (See Table 5.36.) However, when checking the standardised frequencies, it is found that the NTESOL participants used these two types less often than the LIVT and TESOL participants. The NTESOL participants used the first and the second types of use about half as often as the LIVT participants, and about three times less often than the TESOL participants. Moreover, the TESOL participants used the collocation 你说[you say] for referencing what the speaker said in their talk more often than the other participants. (They even used 你说[you say] in this kind of use slightly more often than the LIVT participants.) The TESOL participants' significantly strong tendency to use 你说 [you say] in this way shows our hypothesis 4 is challenged. Even though the TESOL and NTESOL participants had been in very similar language environments outside the classroom, their use of language is still different.

We found that none of the participants used the third type very often. There is only one occurrence of the third type of use in the LIVT data (0.03 times per 1000 words), 3 occurrences of the third type of use in the TESOL data (0.08 times per 1000 words) and 2 occurrences of the third type of use in the NTESOL data (0.05 times per 1000 words.) However, the LIVT group is found to differ from the other two groups in the fourth type of use. In Table 5.36, it is found that the LIVT participants had 6 instances of the fourth type of use of 你说 [you say] (0.16 times per 1000 words.) There was no instance of the fourth type found in the TESOL group and only 1 instance found in the NTESOL group (0.02 times per 1000 words.) What we have found proves our null hypothesis 3 is incorrect. The LIVT participants had a comparatively strong tendency to use the word 说 [to say] as an intensifier/emphasiser. It was a discourse difference but also, to some degree, a lexico-grammatical difference too. Whether this difference can be seen as an effect of the LIVT participants' different English input requires more research. Nevertheless, what we have shown is that there was a difference between the LIVT participants and the participants in the other two groups in their spoken Mandarin.

5.3.3 Summary of the use of 你 (nǐ) [you]

In our investigation on the use of 你(nǐ)[you], we have found several noticeable differences between the three groups. Therefore, our null hypotheses are found to be incorrect. The findings of our investigation into the use of the collocations of 你 (nǐ) [you] are presented in Table 5.37 and Table 5.38.

In our examination of the use of ràng 让 n ǐ 你[let you], we found supports for hypothesis 1. The use of *let * to + verb* in the TESOL and NTESOL groups strongly suggests that the LIVT participants used language in a different way from the participants of the other groups. Their comparatively greater exposure to and practice of English may be the reason for the LIVT group not using this pattern. There was another lexico-grammatical difference between the LIVT group and the TESOL and NTESOL groups. The LIVT participants used the pattern verb + n ǐ 你 shuō 说NP/[S] (shuō 说 as an emphasiser 'saying') more often than the TESOL and NTESOL participants. Although we can still find instances of this pattern in the TESOL and NTESOL groups, what we found is that the LIVT participants were primed for the pattern comparatively more strongly than the TESOL and NTESOL participants were. Whether this stronger priming was influenced by the LIVT participants' extra English input needs further study, but the finding shows that the LIVT participants used Mandarin differently from those in Taiwan in some respects. All the noticeable differences listed in Table 5.37 show that our null hypothesis 3 is incorrect.These differences between the TESOL and NTESOL groups appear to be primarily discourse in nature. However, the finding of these noticeable differences between the TESOL and NTESOL groups indicates our hypothesis 4 is incorrect. Next, we will investigate the use of impersonal pronoun t ā 它 [it] in the three groups.

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	**Collocations**	**The type of differences**	**Description OR noteworthy findings/observation**
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	ràng 让 nǐ 你 [let you]	Discourse preference (Also, different lexico-grammatical use in the use of *let * to verb* between the LIVT and the TESOL and NTESOL groups – Hypothesis 1)	The LIVT participants tended to use the word qù 去 [to go to] for adding a sense of activeness when they used the word ràng 让 [let] in their Mandarin. (Discourse preference) From observation on the patterns ràng 让 * qù 去 + verb [let * to go to + verb] and *let * to +verb*, the LIVT participants were different in the use of English from the TESOL and NTESOL participants.
The LIVT participants used the collocations **significantly more frequently** than the NTESOL participants.	shíhou 时候 nǐ 你 [(time) you]	Discourse preference	The LIVT participants made mention of the time or condition more frequently than the NTESOL participants, especially when the subject was the second person nǐ 你[you]. (Discourse preference) The LIVT participants showed similar discourse preferences in both their English and Mandarin use of *when you*/de 的shí 时hou 候nǐ 你 [of time/when you]. (Discourse preference)
	nǐ 你 yào 要 [you want]	Discourse preference	The LIVT participants discussed more about what a person should do, and the NTESOL participants talked about what a person would want to do. The LIVT and TESOL groups and the NTESOL group had their own preferences in using the collocations yào 要 (nǐ 你yào 要) [you want/must/going to]. (Discourse preference)
	nǐ 你 shuō 说 [you say]	Discourse preference / Lexico-grammatical difference	The LIVT participants tended to use the word shuō 说 [to say] as an intensifier/emphasiser more often than the NTESOL participants.] (Discourse preference) The LIVT participants used the pattern *verb* + nǐ 你 shuō 说 + *NP/[S]*(shuō 说 as an emphasiser 'saying') more often than the TESOL and NTESOL participants. This kind of use was not found in the TESOL group. (Lexico-grammatical difference)

Table 5.37 Findings in the use of 你(nǐ)[you] that Disagree with Hypothesis 3

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	**Collocations**	**The type of differences**	**Description OR noteworthy finding/observation**
The TESOL participants used the collocations **significantly more frequently** than the NTESOL participants.	shí hou 时候 nǐ 你 [(time) you]	Discourse preference	The TESOL participants provided description of time or condition more frequently than the NTESOL participants, especially when the subject was the second person nǐ 你[you]. (Discourse preference)
	nǐ yào 你要 [you want]	Discourse preference	The TESOL participants discussed more what a person should do, and the NTESOL participants talked about what a person would want to do. The LIVT and TESOL participants and the NTESOL participants had their own preferences in using the collocations yào 要 (nǐ yào 你要) [you want/must/going to]. (Discourse preference)
	nǐ shuō 你说 [you say]	Discourse preference	The TESOL participants used the collocation nǐ 你 shuō 说 [you say] for referencing what the other speaker said in their talk more often than the other groups. (Discourse preference)

Table 5.38 Findings in the use of 你(nǐ)[you] that Disagree with Hypothesis 4

5.4 The use of pronoun 它 (tā) [it]

5.4.1 L1 Collocates of 它 (tā) [it]

From the Top 10 most frequent L1 collocate lists (Table 5.39.), we can see that there are many words shared as L1 collocates of 它 [it] by all the three groups. There are some words, however, appearing only in one group's lists or missing from one group's top 10 most frequent collocate list. The words 它 [it] and 了解 [to understand] appear as L1 collocates only in the LIVT and NTESOL groups' lists. The word 说 [to say] appear only in the LIVT and TESOL groups' lists. The word 对 [to/towards] only shows in the LIVT group's list. The words 看 [to see] and 所以 [so/therefore] appear only in the TESOL group's list. The word 那 [that] only appears in the NTESOL group's list. These findings suggest that there may be some noticeable differences between groups. Next, we will check whether any word frequencies stand out with a view to test our hypotheses 3 and 4.

它	LIVT	331	8.58	TESOL	296	8.38	NTESOL	515	12.23
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	把	22	0.57	把	27	0.76	因为	45	1.07
2	觉得	19	0.49	就是	24	0.68	觉得	42	1.00
3	就是	18	0.47	因为	18	0.51	就是	39	0.93
4	因为	16	0.41	觉得	15	0.42	把	27	0.64
5	有	10	0.26	然后	10	0.28	有	21	0.50
6	知道	9	0.23	有	9	0.25	**它**	19	0.45
7	**它**	9	0.23	**说**	8	0.23	知道	11	0.26
8	**说**	8	0.21	知道	6	0.17	**了解**	11	0.26
9	对	8	0.21	看	6	0.17	是	11	0.26
10	**了解**	7	0.18	所以	6	0.17	那	8	0.19

Table 5.39 L1 Collocates of 它(tā) in LIVT, TESOL and NTESOL

LIVT - NTE						LIVT - TE				TE - NTE					
Less						Less				Less					
Word	With	LLR	Sig.			Word	With	LLR	Sig.	Word	With	LLR	Sig.		
因为	它	11.904	0.001	***	-					**觉得**	**它**	**9.029**	**0.003**	**	-
觉得	**它**	**6.968**	**0.008**	**	-					因为	它	7.712	0.005	**	-
就是	它	6.169	0.013	*	-					它	它	6.442	0.011	*	-
										了解	**它**	5.384	0.020	*	-

Table 5.40 Log-likelihood Ratios for L1 collocates of 它(tā) in LIVT, TESOL and NTESOL

In Table 5.40, it is found that the LIVT participants used the collocation 就是 它 [just it] less frequently than the NTESOL participants (with weak significance). In 5.2.1, we found that the LIVT participants used the collocation 就是 我 [just I] significantly more often than the other participants. We also found that the LIVT participants used 就是/就[just] more often than the other participantss. The adverb 就是 [just] is often used in L1 position of the pronoun. The finding that the LIVT participants used the collocation 就是 它 [just it] less frequently than the NTESOL participants might be explained by the fact that the LIVT participants used the pronoun 它 [it] significantly less often than the NTESOL participants, and the LIVT participants did not use 就是 [just] as often as the NTESOL participants did when the subject was an impersonal pronoun 它 [it]. This finding challenges our hypothesis 3, in which we supposed there would be no noticeable difference in their use of Mandarin between those Taiwanese who were studying in the UK and those who were studying in Taiwan. What we have found is that the different groups had their own discourse preferences when using certain combinations of words.

In Table 5.32, the TESOL participants used the collocation 它 它 [it it] and 了解 它[to understand it] significantly less frequently than the NTESOL participants. These two findings challenge our hypothesis 4, where we hypothesised that Taiwanese students studying in English-relevant subjects would show no noticeable differences in their Mandarin from those studying in non-English-relevant subjects. Despite of the use of repetition (stammering) 它 它[it it], which we will not investigate in the current study, the significant difference in the use of 了解 它 [to

understand it] between the TESOL and NTESOL groups is worth noting. We will investigate the use of 了解 它 [to understand it] later. We also see that the LIVT and TESOL participants used the collocations 因为 它 [because it] and 觉得 它 [to think it] less frequently than the NTESOL participants. These two findings also challenge our hypotheses 3 and 4. As we have mentioned in our previous discussion, the use of Mandarin connectives wii not be included in this study. We will however investigate the collocation 觉得 它 [to think it] in detail to test our hypotheses 3 and 4.

觉得(juéde) 它(tā) [to think it]

The LIVT and TESOL participants used the collocation 觉得 它 [to think it] significantly less frequently than the NTESOL participants. The NTESOL participants used the collocation 觉得 它 [to think it] once per 1000 words. The LIVT participants used 觉得 它(tā) [to think it] 0.49 times per 1000 words, and the TESOL participants used 觉得 它 [to think it] 0.42 times per 1000 words. The LIVT and TESOL participants used the collocation 觉得 它 [to think it] twice less often than the NTESOL participants. (See Table 5.41.)

	LIVT				**TE**				**NTE**			
	Freq.		%	Per 1000 words	Freq.		%	Per 1000 words	Freq.		%	Per 1000 words
觉得 (juéde)	685		100	17.76	559		100	15.83	781		100	18.54
觉得 它(juéde tā)	19		2.77	**0.49**	15		2.68	**0.42**	42		5.38	**1.00**
Patterns of 觉得 它 (juéde tā) (TOP 5)	N	L1	Centre	R1	N	L1	Centre	R1	N	L1	Centre	R1
	1	我	觉得它		1	我	觉得它		1	我	觉得它	是
		wǒ	juédé tā			wǒ	juédé tā			wǒ	juédé tā	shì
		I	think it			I	think it			I	think it	to be

Table 5.41 Use of 觉得 (juéde) and 觉得 (juéde) 它(tā) in LIVT, TESOL and NTESOL

When investigating how this collocation was used by the three groups, it is found that its most common use is in the pattern 我 觉得 它 (是) [I think it (is)]. There are 16 occurrences of the

collocation我 觉得 它 [I think it] used by the LIVT participants (0.41 times per 1000 words), and it accounts for 84.21% of the total use of 觉得它 [think it]. There are 15 occurrences of the use of 我 觉得 它[I think it] used by the TESOL participants (0.41 times per 1000 words), and it accounts for 80% of the total use of 觉得 它 [think it]. There were 31 occurrences of the collocation 我 觉得 它 [I think it] (0.74 times per 100 words) used by the NTESOL participants, and it accounts for 73.81% of the total use of 觉得它 [think it]. (See Table 5.42)

		LIVT			**TE**			**NTE**		
		Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
	我(wǒ) 觉得 (juéde)	487	--	**12.63**	372	--	**10.53**	519	--	**12.32**
	觉得 (juéde) 它(juéde tā)	19	100	0.49	15	100	0.42	42	100	1.00
1	我(wǒ) 觉得 它(juéde tā)	16	84.21	0.41	12	80.00	0.34	31	73.81	0.74
2	Omitting 我 (wǒ)	1	5.26	0.03	0	0.00	0.00	7	16.67	0.17
3	我(wǒ) + adv/modal verb/neg. marker+觉得 (juéde)	2	10.53	0.05	2	13.33	0.06	2	4.76	0.05
4	Other subjects (e.g. you, the children)	0	0.00	0.00	1	6.67	0.03	2	4.76	0.05

Table 5.42 Use of 我 觉得 (juéde) and 我 觉得 (juéde) 它 (tā) in LIVT, TESOL and NTESOL

This finding echoes what we have found with regard to the use of 我 觉得 我 [I think I]. The verb 觉得 [think] in the collocation 觉得 它 [to think it] is a part of the collocation 我 觉得 [I think], which serves mainly as a discourse strategy to start a sentence. Although the pattern 我 觉得 它 [I think it] is the primary use in the use of the collocation 觉得 它 [think it] [Use 1] in all the three groups, it is found that there are some instances in which the collocation 觉得 它 [think it] is not following the pronoun 我 [I] directly. We can categories these instances into three kinds of use. The first kind is the collocation 觉得它 [think it] whose use is similar to the pattern 我 觉得 它

[I think it] as a sentence starter, except that the pronoun 我 [I] is omitted. (See Example A in Table 5.43.) The omission of pronouns/nouns understood by both the speaker and the hearer is very common in Mandarin. There are 7 occurrences of the first kind of uses (我) 觉得 它 [(I) think it] in the NTESOL group. There is only one occurrence of this use in the LIVT groupo and no such use at all in the TESOL group. These findings prove our null hypotheses 3 and 4 are incorrect. The LIVT and TESOL participants used fewer instances of omission of the pronoun 我 [I] in this context than the NTESOL participants. The second kind of uses of 觉得 它 is a variant form of the pattern 我 觉得 它 [I think it]. This kind of use has an adverb, modal verb or negative markers between the words 我 (I) and 觉得它 (think it.) The use of 我 * 觉得 它 [I * think it] is not a discourse strategy for buying time, but has a literal meaning. (See Example B in Table 4.43.) The three groups did not show great differences with regard to this use. There are 2 instances of this use in all the three groups. The standardised frequencies suggest that the participants of the three groups used this kind of use with similar frequencies. The third kind of use does not make use of 我[I]. It has other nouns/pronouns as the sentence subject. Similar to the previous use, it is not a discourse strategy for buying time, but again has a literal meaning. (See Example C in Table 4.43.) This use was rarely found in these groups. There were 2 occurrences of this kind of use in the NTESOL data, only 1 occurrence of this use in the TESOL data and none in the LIVT data.

1	[我] 蛮 喜 欢 文 言 文 啊 , 因 为 [我] 觉得 它 蛮 美 的
	[Wǒ] mán xǐhuɑn wényánwén ā , yīnwèi [wǒ] juédé tā mán měi de
	[I] am quite fond of classic prose, because [I] think it is quite beautiful.
2	我 没 有 觉得 它 对 你 [有 帮 助 啦]
	Wǒ méi yǒu juédé tā duì nǐ [yǒu bāng zhù lā]
	I don’t think it will be helpful to you.
3	小 孩子 也 特别的是 , 欸 , 能 够 觉得 它 饶富 趣味
	Xiǎoháizi yě tèbié de shì , èi , nénggòu juédé tā ráofù qùwèi
	The children too, in particular, can think it’s full of fun.

Table 5.43 Examples of another three types of the use of 觉得 (juéde) 它 (tā) in LIVT, TESOL and NTESOL

Here we have 3 noticeable findings: 1) the LIVT and TESOL participants did not use 觉得 [think] with the pronoun 它 [it] as frequently as the NTESOL participants, but 2) when they used the collocation 觉得 它 [think it], their primary use/collocation (我觉得它 [I think it]) was the same as

the NTESOL participants. However, 3) the NTESOL participants omitted the pronoun 我 [I] in the discourse use of the pattern 我 觉得 它 [I think it] more often than the LIVT and TESOL participants. When the LIVT and TESOL participants used 觉得 [think] with 它 [it] in the sense of 我觉得它 [I think it], they did not omit the pronoun 我 [I] as frequently as the NTESOL participants. What we have found shows the LIVT and TESOL participants sometimes behaved quite differently from the NTESOL participants in their Mandarin use. Therefore, our null hypotheses are incorrect. It is difficult to see why the NTESOL participants and the LIVT and TESOL participants behaved differently in their use of 觉得 它 [think it] (and the word 它 [it] since it is twice more often used by the NTESOL participants than by the LIVT and TESOL participants. It is possible that the NTESOL participants or the LIVT and TESOL participants had different primings for the use of the word 它 [it] , and more investigation will be needed in the future.

了解(liǎojiě) 它(tā) [to understand it]

The TESOL participants used the collocation 了解 它 [understand it] significantly less often than the NTESOL participants. The frequency of the use of 了解 它 [understand it] in the LIVT group does not have significant differences from the frequencies of this use in the TESOL and NTESOL groups. The NTESOL participants used the collocation 了解 它 [understand it] 0.26 times per 1000 words. The TESOL participants used the collocation 了解 它 [understand it] 0.06 times per 1000 words. The LIVT participants used this collocation 0.18 times per 1000 words. (See Table 5.46.) The TESOL participants used the collocation 了解 它 [understand it] four times less often than the NTESOL participants, (and half as often as the LIVT participants.) This finding suggests that our hypothesis 4 is challenged. The participants studying in English-relevant subjects had behaved differently from those studying in non-English-relevant subjects in terms of the frequencies of using the collocation 了解 它 [understand it]. Initially, we thought the reason behind the different frequencies between these groups might be discourse in nature. The TESOL participants might not talk about the comprehension of something as often as the NTESOL participants. However, when we checked the use of the word 了解 [understand] and the collocation 了解 它 [understand it] in

the three groups, it is found that the TESOL participants actually used the word 了解 [understand] more often than the NTESOL participants, and the way the collocation 了解 它 [understand it] was used by the TESOL participants differed from the way by the NTESOL participants (and also by the LIVT participants.)

In Table 5.44, we can see that the TESOL participants used the word 了解 [understand] 37 times (1.05 times per 1000 words) and the collocation 了解 它 [understand it] twice (0.06 times per 1000 words.) The use of 了解 它 [understand it] accounts for only 5.41% of the total use of 了解 [understand] in the TESOL group. The NTESOL participants, however, used the word 了解 [understand] 31 times (0.74 times per 1000 words) and the collocation 11 times (0.26 times per 1000 words.) The use of the collocation 了解 它 [understand it] accounts for 35.48% of the total usage of 了解 [understand] in the NTESOL group. The LIVT participants used 了解 [understand] 35 times (0.91 times per 1000 words) and the collocation 了解 它 [understand it] 7 times (0.18 times per 1000 words). The use of 了解 它 [understand it] accounts for 20% of the total use of 了解 [understand] in the LIVT group. What we have found suggests that even though the TESOL participants used the verb 了解 [understand] slightly more often than the other participants, the TESOL participants tended not to use 它[it] as the object when they used the verb 了解 [understand]. (See the patterns in Table 5.44.) Taking the L1 and R1 collocates of the words 了解 它 [understand it] into our comparison, and we find that the TESOL participants did not even share similar collocates. The LIVT and NTESOL participants had the word 去 [go to] as a L1 collocate of 了解 它 [understand it] and the word 的 [~'s/of] as a R1 collocate of 了解 它 [understand it] in common. The TESOL participants had neither of these words as collocates of 了解 它 [understand it]. What we find here suggests that the TESOL participants used this collocation differently from the other participants (the NTESOL participants in particular.) Therefore, our hypothesis 4 is shown to be incorrect.

	LIVT			TE			NTE		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
了解(liǎojiě)	35	100	0.91	37	100	**1.05**	31	100	**0.74**
了解(liǎojiě)它(tā)	7	**20**	0.18	2	**5.41**	**0.06**	11	**35.48**	**0.26**
Patterns of 了解(liǎojiě) (TOP 5)	N L1 Centre R1 1 可以 了解 它 kéyǐ liáojiě tā can understand it 2 去 那 qù nà go to that 3 不 然后 bú ránhòu not then 4 要 yào want/must			N L1 Centre R1 1 去 了解 那 qù liáojiě nà go to understand that 2 的 说 de shuō of (DE) say 3 可以 对 kéyǐ duì can to/towards 4 就是 jiùshì just			N L1 Centre R1 1 去 了解 它 qù liáojiě tā go to understand it 2 可以 kéyǐ can		
L1 collocate of 了解(liǎojiě)它(tā)	去[go to] 2 要[want] 1 会[can] 1			希望[hope] 1 为了[for] 1			可以[can] 3 去[go to] 3 就是[just] 1		
R1 collocate of 了解(liǎojiě)它(tā)	的[~'s/of] 2 然后[then] 1 意思[meaning] 1			里面[inside] 1 是[to be] 1			**的[~'s/of] 5** 在[at/in] 1		

Table 5.44 Use of 了解(liǎojiě) and 了解(liǎojiě)它(tā) in LIVT, TESOL and NTESOL

The finding also echoes what we have found concerning the use of English *it* in the TESOL group. With regard to the use of *use it* and *learn it* (in 4.4.1), we found that the TESOL participants did not use the pronoun *it* as the object of the verb *learn* and *use* as often as the participants in the other two groups did (much less often than the NTESOL participants in particular.) This consistency in the use of English *it* and the Mandarin 它[it] in the TESOL group suggests that, perhaps, the TESOL participants' primings for the English impersonal pronouns *it* and Mandarin 它[it] had been shared in some way, even though they were from two different languages. It is similar to the consistent use of *I just*/我就 (就是) [I just] and *I think*/我觉得[I think] that we found in the LIVT group. The adverbs *just* and 就 [just] and the verbs *think* and 觉得[think] are alike in the LIVT group's language use. Here, we found possible evidence for the existence of cross-language primings for words and collocations that people have when they are exposed to the two languages at the same time.

5.4.2 R1 Collocates of 它 (tā) [it]

From the top 10 most frequent R1 collocate lists (Table 5.45), we can see that there are some R1 collocates shared by all the three groups. Those words are 的 [of/~'s], 是 [be], 可能 [maybe], 有 [have], and 就 [just]. There are also some collocates that only appear in one group's list or are missing from one group's list. The words 会 [can] and 其实 [actually] are missing from the LIVT group's list. The words 就是 [just], 它 [it] and 也 [as well] are missing from the TESOL group's list. The words 对 and 那 only appear in the LIVT group's list. The words 都 [both/all], 可以 [can] and 到底 [at all/finally] only appear in the TESOL group's list. We now check whether there are any noticeable differences between groups in terms of their frequencies.

它	LIVT	331	8.58	TESOL	296	8.58	NTESOL	515	12.23
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	的	44	1.14	的	38	1.08	的	53	1.26
2	**就是**	20	0.52	是	23	0.65	是	38	0.90
3	是	19	0.49	**会**	16	0.45	**它**	14	0.36
4	可能	13	0.34	有	12	0.34	**就是**	12	0.33
5	有	12	0.31	就	10	0.28	有	11	0.28
6	**也**	11	0.29	可能	10	0.28	可能	11	0.26
7	对	11	0.29	都	8	0.23	**其实**	11	0.26
8	那	10	0.26	**其实**	7	0.20	**会**	10	0.26
9	**它**	9	0.23	可以	6	0.20	**也**	10	0.24
10	就	8	0.21	到底	6	0.17	就	9	0.24

Table 5.45 R1 Collocates of 它(tā) in LIVT, TESOL and NTESOL

When checking for significance (see Table 5.46), it is found that the LIVT participants used the collocation 它 那 [it that] significantly more often than the TESOL and NTESOL participants. The LIVT participants used the collocations 它 就是 [it just] and 它 对 [it to/towards] significantly more often than the TESOL participants, and the collocation 它 会 [it can] significantly less often than the TESOL participants. The TESOL participants used the collocation 它 都 [it both/all]

significantly more often than the LIVT and NTESOL participants. The TESOL participants used the collocation 它 会 [it can] significantly more often than the LIVT participants and the collocations 它 它 [it it], 它 就是 [it just] and 它 的 [it ~'s/of; its] significantly less frequently than the NTESOL participants. Both the LIVT and TESOL participants used the collocation 它 是 [it *BE*] significantly less frequently than the NTESOL participants. These differences challenge our null hypotheses 3 and 4. As mentioned previously, we will not be investigating the use of repetition (它它[it it]) and the modal verb (它 会[it can]) in the current study. The collocation 它 是 [it *BE*] will be investigated in chapter 6.

LIVT - NTE						LIVT - TE						TE - NTE					
Over						Over						Over					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
它	**那**	**4.048**	**0.044**	*	+	**它**	**那**	**8.912**	**0.003**	**	+	它	都	4.981	0.026	*	+
						它	就是	7.546	0.006	**	+						
						它	对	4.197	0.041	*	+						
Less						Less						Less					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
它	是	12.482	0.000	***	-	它	会	4.436	0.035	*	-	它	它	9.028	0.003	**	-
						它	都	4.390	0.036	*	-	它	是	6.530	0.011	*	-
												它	的	4.506	0.034	*	-
												它	就是	4.295	0.038	*	-

Table 5.46 Log-likelihood Ratios for R1 collocates of 它(tā) in LIVT, TESOL and NTESOL

The use of the adverb 就是 [just] has been investigated, so we will not consider the collocation 它 就是[it just] further. The use of 它 的 [it ~'s/of; its] is similar to the English possessive *its* and the collocation and colligation of 它 的[it ~'s/of] is comparatively simple without there being much room for investigation into lexico-grammatical differences. Therefore, we will not take the collocation 它的 [it ~'s/of] for further investigation. The rest of the collocations are 它 那 [it that], 它对 [it to/towards] and 它都 [it both/all]. In the use of 它都 [it both/all], it is found that there are no noticeable differences between the groups in terms of the lexico-grammatical features of the word combination.. The frequent use of 它都 [it both/all] in the TESOL group indicates that the

TESOL participants preferred to use the adverb 都 for emphasis more frequently than the LIVT and NTESOL participants. We will, therefore, not include the collocation 它都 [it both/all] in our further investigation. Since our aim is to find whether there is anything worthy of note that differentiates the LIVT group (participants studying in the UK) and the TESOL group (participants studying in English-relevant subjects) from the NTESOL group, the collocation 它 那 will be better than the collocation 它对 [it to/towards] for testing our null hypotheses 3 and 4. The LIVT participants used the collocation 它那 [it that] significantly more often than the TESOL and NTESOL participants.

它 那 (tā nà) [it that]

The LIVT participants used the collocation 它那 [it that] significantly more often than the TESOL and NTESOL participants. The collocation 它那 [it that] was used 10 times by the LIVT participants (0.26 times per 1000 words), once by the TESOL participants (0.03 times per 1000 words) and 4 times by the NTESOL participants (0.09 times per 1000 words.) The LIVT participants used the collocation 它那 [it that] almost 9 times more frequently than the TESOL participants and almost 3 times more often than the NTESOL participants. (See Table 5.47.)

		Freq.	Per 1000 words		LLR	Sig.		
它 那 (tā nà)	LIVT	10	0.26	**LIVT-TE**	8.912	**0.003**	**	+
	TE	1	0.03	**LIVT-NTE**	4.048	**0.044**	*	+
	NTE	4	0.09	TE-NTE	2.882	0.090		-

Table 5.47 Frequency and Log-likelihood Ratios: 它 那 (tā nà) in LIVT, TESOL and NTESOL

The word 那 [that] can function as a demonstrative, which is used to refer to an item. (e.g. 那是我的书 = That is my book.) The word 那 [that] can also function as a specifier, which is used as a part of a noun phrase to identify a specific item (e.g. 那本书 = that book) (Ross and Ma, 2006, 2014). When we observe all the instances of the use of 它那 [it that], it is found that the word 那

[that] in the context of 它那 [it that] functions as a specifier. The pronoun 它 [it] is used for adding emphasis. The examples in Table 5.48 show the word 那 [that] being used as a part of a noun phrase ('that meaning'/ 'that feeling'), and the word 它 [it] is used to point out what this noun phrase relates to 'that meaning of it (the word)' / 'that feeling of it (the language). The word 它 [it] in this usage functions similarly to an English possessive ('its meaning' / 'its feeling'). We can substitute the word 它 [it] into the Mandarin possessive 它[it] 的 [~'s/of] without affecting the meaning of sentences.

1	就 是 要 用 不同 的 字 去表达 **它 那** 个 意思 =**它 的 那** 个 意思
	jiùshì yào yòng bùtóng de zì qù biǎodá tā nà gè yìsi
	[you] just need to use different words to convey **that** meaning (**of it** = the word; the meaning of the word (it)).
2	学 一 个 新 的 语言 然后 你 要 翻 翻 译 **它 那** 种 感 觉 = **它 的 那** 种 感 觉
	xué yī gè xīn de yǔyán ránhòu nǐ yào fān fānyì tā nà zhǒng gǎnjué
	[you] learn a new language, and then you have to translate **that** feeling (**of it** = the language).

Table 5.48 Examples of the use of 它 那 (tā nà)

What we found in the use of 它那 [it that] in the LIVT group shows that the LIVT participants used this kind of expressions much more often than the TESOL and NTESOL participants. When we further checked how many instances of the pattern *pronouns with specifier 那 [that]* occurred in all the three groups, we found that the LIVT participants used the pattern *pronoun + 那 [(of) it, that]* more often than the TESOL and NTESOL participants. The LIVT participants used the pattern *pronoun + 那 [(of) it, that]* 44 times (1.06 times per 1000 words.) The TESOL participants used this pattern 22 times (0.62 times per 1000 words), and the NTESOL participants used this pattern also 22 times (0.52 times per 1000 words.) (See Table 5.49.) The TESOL and NTESOL groups are alike in terms of the frequencies of this pattern, and the LIVT group shows a higher frequency of the pattern.

	LIVT			TE			NTE		
	Freq	%	Per 1000 words	Freq	%	Per 1000 words	Freq	%	Per 1000 words
那 [that]	901	100	23.36	752	100	21.29	736	100	17.47
它 那 [(of) it, that]	10	1.11	0.26	1	0.13	0.03	4	0.54	0.09
它+ 的 +那 [of it, that]	2	0.22	0.05	2	0.27	0.06	1	0.14	0.02
Pronoun+ 那 [(of) it, that]	41	4.55	**1.06**	22	2.93	0.62	22	2.99	0.52
Pronoun + 的 +那 [of it, that]	8	0.89	**0.21**	8	1.06	**0.23**	4	0.54	0.09

Table 5.49 Use of the patterns *pronoun+ 那(nà) [(of) it, that]* / *pronoun + 的(de) +那(nà) [of it, that]*

We also checked the pattern where the Mandarin possessive marker 的 was used. The pattern *pronoun + 的 + 那 [of it, that]* was used 8 times by the LIVT participants (0.21 times per 1000 words), also 8 times by the TESOL participants (0.23 times per 1000 words) and 4 times by the NTESOL participants (0.09 times per 1000 words.) With regard to the pattern *pronoun + 的 + 那 [of it, that]*, it is found that the LIVT and TESOL groups are alike, and the NTESOL group has a lower frequency of this pattern. These findings further demonstrate that the null hypotheses are incorrect. The LIVT and the TESOL participants had a very different use of this specific kind of language use from that of the control group (NTESOL participants). If we see the pattern *pronoun + 那 [(of) it, that]* as the default use (because of its higher use), the LIVT participants showed a stronger tendency to use such a pattern than the participants in Taiwan. Whether this stronger tendency is caused by their different language learning backgrounds in an English-speaking country is not clear. However, from oberservation on the use of the second pattern *pronoun + 的 + 那 [of it, that]*, we may be able to see how the English input might possibly cause a difference in language use between the groups. The LIVT and TESOL participants showed a similar frequency of the use of the Mandarin possessive marker 的 [~'s/of] in a pattern where the word 的 [~'s/of] is usually missed. In English, the possessive has its own form and cannot be omitted in a noun phrase. The LIVT and TESOL participants, who had been exposed to English more than the NTESOL participants had, may

be influenced by their English primings and therefore had a stronger tendency to use the possessive 的 [~'s/of] in a context where normally it was not used in Mandarin.

5.4.3 Summary of the use of pronoun 它(tā) [it]

In our investigation on the use of 它 (tā) [it], we have found several noticeable differences between the three groups. Therefore, our null hypotheses were found to be incorrect. The findings of our investigation are presented in Table 5.50 and Table 5.51. With regard to our test of null hypothesis 3, we found that the LIVT participants behaved differently from the TESOL and NTESOL participants. The LIVT participants did not use the discourse starter 我 觉得 [I think] with 它 [it] as often as the NTESOL participants. They also showed less use of the omission of 我 [I] in the expression 我 觉得 它 [I think it]. The omission of 我 [I] is a common discourse use in Mandarin. When a speaker thinks there is no need to spell out who is doing the action or who is involved, s/he tends to omit the pronoun. The LIVT participants used this discourse use less frequently than the other participants.

The LIVT participants used the pronoun 它 [it] and other pronouns with the specifier 那 [that] more often than the TESOL and NTESOL participants. The pronouns in this context are usually used to point out what the noun phrase (containing the specifier 那 [that]) refers to. It is a discourse usage for coherence. In this kind of use, however, we found one lexico-grammatical difference between the LIVT group and NTESOL group. The possessive marker 的[~'s/of] can be used in the pattern *pronoun + 那(nà)* [(of) it, that]. There were more instances of the use of the pattern *pronoun+ 的 + 那(nà)* [of it, that] found in the LIVT data. This pattern was less used by the NTESOL participants. The strength of use of 的[~'s/of] in the LIVT group suggests that the LIVT participants might be primed to use 的[~'s/of] differently from the NTESOL participants. There is a possibility that their different English input might have some influence, but this needs more research. Nevertheless, what we have found in the investigation shows that null hypothesis 3 is incorrect.

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	Collocations	The type of differences	The possible factors which affected this difference OR noteworthy observation
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	它 那 [it that]	Discourse preference / Lexico-grammatical difference	The LIVT participants used the word 它 [it] with the specifier那 [that] for referring the item to which the noun phrase was related more often than the TESOL and NTESOL participants. (Discourse preference) The LIVT participants used the pattern *pronoun + 那 [(of) it, that]* more often than the TESOL and NTESOL participants (Discourse preference) The LIVT and TESOL participants showed similar frequencies of the use of Mandarin possessive marker 的 [~'s/of] in the pattern where the word 的 [~'s/of] is usually missed. This pattern was relatively used less by the NTESOL participants. (Lexico-grammatical difference.)
The LIVT participants used the collocations **significantly less frequently** than the NTESOL participants.	觉得它 [to think it]	Discourse preference	The LIVT participants did not use 觉得 [think] with the pronoun 它 [it] as frequently as the NTESOL participants. (The use of觉得 [think] with the pronoun 它 [it] serves a discourse function mostly.) (Discourse preference.) The LIVT participants did not omit the pronoun 我 [I] in the discourse use of the pattern 我觉得它[I think it] as often as the NTESOL participants. (Discourse difference)

Table 5.50 Findings in the use of 它(tā) [it] that disagree with Hypothesis 3

With regard to null hypothesis 4, we found that the TESOL participants also did not use the discourse starter 我 觉得 [I think] with 它 [it] as often as the NTESOL participants did. They also less often omitted 我 [I] in the use of 我 觉得 它 [I think it].

Findings that Disagree with Hypothesis 4 (null hypothesis)			
Noticeable Differences	Collocations	The type of differences	The possible factors that affected this difference OR noteworthy observation
The TESOL participants used the collocations **significantly less frequently** than the NTESOL participants.	觉得它 [to think it]	Discourse preference and lexico-grammatical difference	The TESOL participants did not use 觉得 [think] with the pronoun 它 [it] as frequently as the NTESOL participants. The use of 觉得 [think] with the pronoun 它 [it] serves a discourse function mostly. (Discourse preference.) The TESOL participants did not omit the pronoun 我(wǒ) [I] in the pattern 我觉得它 [I think it] as often as the NTESOL participants. (Lexico-grammatical difference)
	了解它 [to understand it]	Discourse preference and lexico-grammatical difference	The TESOL participants did not use 它 as the object when they used the verb 了解[to understand] as much as the NTESOL participants. (Discourse preference/lexico-grammatical difference) The TESOL participants did not share similar collocates of 了解它 with the LIVT and NTESOL participants. (lexico-grammatical difference)
The TESOL participants used the collocations **significantly less frequently** than the LIVT participants.	它 那 [it that]	Discourse preference and lexico-grammatical difference	The TESOL participants did not use the word 它 [it] with a noun phrase containing the specifier 那 [that] for adding a link to the item the noun phrase refers as often as the LIVT participants. (Discourse difference) The LIVT and TESOL participants showed similar frequencies in the use of Mandarin possessive marker 的[~'s/of], a pattern where the word 的[~'s/of] is usually missed. The fuller pattern was used less by the NTESOL participants. (Lexico-grammatical difference)

Table 5.51 Findings in the use of 它 (tā) [it] that disagree with Hypothesis 4

The TESOL participants did not use the pronoun 它[it] and other pronouns with the specifier 那[that] as often as the LIVT participants. Similarly, the TESOL participants used the possessive marker 的[~'s/of] in the pattern *pronoun + 那(nà)* [(of) it, that] more often than the NTESOL participants. This suggests that the TESOL participants were also more strongly prone to use 的[~'s/of] than the NTESOL participants. There is one finding with regard to the use of 了解 [understand] and 它 [it] that suggests the TESOL participants were different in the use of 它 [it] from the NTESOL participants. The TESOL participants did not use the pronoun 它[it] as the object of the verb 了解 [understand] as often as the NTESOL participants. This may be a difference of discourse preference (they did not talk much about what needed to be understood in their discourse) or of lexico-grammatical difference (they did not use the pronoun 它[it] as the object as often as the NTESOL participants did.) These findings suggest that there is a possibility that their different English input might have some influence on their language choices. But further research is required to test this possibility. Nevertheless, what we have found in our investigation shows the hypothesis 4 is incorrect.

5.5 Conclusion

In this chapter, we have investigated the L1 and R1 collocates of Mandarin pronouns 我[I], 你[you] and 它[it]. This investigation is relevant to our hypotheses 3 and 4 – the null hypotheses:

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

We identified several collocations with noticeable differences in their frequencies across the three groups. We also investigated the patterns of these collocations and categorised the use of these collocations according to the different features found in the contexts of these collocations. We hypothesised that there should be little difference in these participants' use of Mandarin. The participants had received similar Mandarin input. They were born, raised and educated in Taiwan. They also had lived in a Mandarin-speaking community for at least two decades. With respect to the use of common and frequent words for which these participants had received abundant input, it is difficult to imagine that their use of such words would be very different, still less that any differences would cluster in any particular group. However, on the basis of what we have seen in the use of the collocations, hypotheses (3 and 4) are incorrect.

In Table 5.52, we can see that the participants did not differ in the use of Mandarin words greatly in terms of their collocations and lexico-grammatical use. What made the frequencies of use of these collocations significantly different between groups is mostly the different discourse preferences of the three groups. The LIVT participants showed a strong tendency to use words that can serve to add extra emphasis or to downtone, or as a strategy for earning more time to generate their sentences. These kinds of discourse behaviours were also used by the other groups, but their tendencies to do so were not as strong as that of the LIVT group.

Summary of the findings that Disagree with Hypothesis 3 (null hypothesis) – Discourse Preference	
Collocations	**Descriptions OR noteworthy findings/observation**
就是(jiùshì)我(wǒ) [just I]	The LIVT participants used the pattern 就是(jiùshì) 我(wǒ) 觉得(juéde) [just I think] as a discourse strategy to start their sentences more often than the TESOL and NTESOL participants.
我(wǒ) 就(jiù) [I just]/ 就 (jiù) [just]	The LIVT participants used 我(wǒ) 就(jiù) [I just]/ 就(jiù) [just] for adding different tones to their discourse more often than the TESOL and NTESOL participants. The LIVT participants used 我(wǒ) 就(jiù) [I just] and *I just* more often than the TESOL and NTESOL participants.
我 (wǒ) 还 (hái) [I still/fairly]	The LIVT participants did not use 我(wǒ) 还(hái) [I still/fairly] to downtone when giving a contrary idea in an utterance as often as the NTESOL participants.
让 (ràng) 你 (nǐ) [let you]	The LIVT participants tended to add an extra tone of activeness with the use of the word 去 [to go to] when they used the word 让 (ràng) [let] in their Mandarin.
时候(shíhou) 你(nǐ) [time you]	The LIVT participants tended to provide description of the time or condition more frequently than the NTESOL participants, especially when the subject was the second person 你 (nǐ) [you]. The LIVT participants showed a discourse preference in both their English and Mandarin for *when you* and 的(de)时候(shíhou)你(nǐ) [of time/when you].
你 (nǐ) 要 (yào) [you want]	The LIVT participants tended to discuss what a person should do, and the NTESOL participants tended to talk about what a person would want to do. The LIVT and TESOL participants and the NTESOL participants used the words 要 (yào) (你(nǐ)要(yào)) [you want/must/going to] for different purposes in the discourse.
你 (nǐ)说 (shuō) [you say]	The LIVT participants had a comparatively strong tendency to use the word 说(shuō)[to say] as an intensifier/emphasiser.
它 (tā) 那 (nà) [it that]	The LIVT participants used the combination 它 (tā) [it] with the specifier 那 (nà) [that] to refer to the item to which the noun phrase refers more often than the TESOL and NTESOL participants. The LIVT participants used the pattern *pronoun* + 那 (nà) [(of) it, that] more often than the TESOL and NTESOL participants.
觉得 (juéde) 它 (tā) [to think it]	The LIVT participants did not use 觉得 (juéde) [think] with the pronoun 它 (tā) [it] as frequently as the NTESOL participants. The use of 觉得 (juéde) [think] with the pronoun 它(tā) [it] serves a discourse function mostly. The LIVT participants did not omit the pronoun 我 (wǒ) [I] in the discourse use of the pattern 我 (wǒ)觉得 (juéde) 它 (tā) [I think it] as often as the NTESOL participants.

Table 5.52 Summary of the findings that disagree with Hypothesis 3 – Discourse Preferences

However, we also found that the LIVT participants used words/patterns differently from the other participants when looking at their lexico-grammatical features. The pattern *verb* + *你[nǐ]说[shuō][you say]* + *NP/[S]* with the use of 说[shuō] [to say] as an emphasiser was found relatively more often in the LIVT group, but rarely in the other two groups. The LIVT participants also used 的[de] [~'s/of] in the pattern *pronoun* + *那[nà][that]* (*pronoun* + *的[de]* + *那[nà]*) which the NTESOL participants used comparatively rarely. (See Table 5.53.)

Summary of the findings that disagree with Hypothesis 3 (null hypothesis) – Lexico-grammatical Differences	
Collocations	**Descriptions OR noteworthy findings/observation**
你 (nǐ) 说 (shuō) [you say]	The LIVT participants used the pattern *verb + 你 (nǐ)说 (shuō)+ NP/[S]* (说 (shuō) as an emphasiser 'saying') more often than the TESOL and NTESOL participants. This use was not found in the TESOL group. (Lexico-grammatical difference)
它 (tā) 那 (nà) [it that]	The LIVT and TESOL participants showed similar frequencies in using the Mandarin possessive marker 的(de)[~'s/of] in the pattern *pronoun* +的(de) +那(nà) where the word 的 (de) [~'s/of] is usually omitted. This pattern was used less by the NTESOL participants. (lexico-grammatical difference.)

Table 5.53 Summary of the findings that disagree with Hypothesis 3 –Lexico-grammatical Differences

Summary of the findings that Disagree with Hypothesis 4 (null hypothesis) – Discourse Preference	
Collocations	**Descriptions OR noteworthy findings/observation**
我(wǒ)就(jiù)[I just]/ 就(jiù) [just]	The TESOL participants did not use 我 (wǒ) 就 (jiù) [I just]/ 就 (jiù) [just] to emphasise or downtone in their discourse as often as the LIVT participants.
我(wǒ) 还(hái) [I still/fairly]	The TESOL participants did not use 我 (wǒ) 还 (hái) [I still/fairly] to downtone when giving a contrary idea in an utterance as often as the NTESOL participants.
时候(shíhou) 你 (nǐ) [time you]	The TESOL participants tended to provide description of the time or condition more frequently than the NTESOL participants, especially when the subject was the second person 你 (nǐ) [you].
你(nǐ)要(yào)[you want]	The TESOL participants tended to discuss more often what a person should do, and the NTESOL participants tended to talk about what a person would want to do. The LIVT and TESOL participants and the NTESOL participants had their own discourse tendencies in their use of 要 (yào) (你 (nǐ) 要 (yào)) [you want/must/*BE* going to].
你 (nǐ) 说 (shuō) [you say]	The TESOL participants used the collocation 你 (nǐ) 说 (shuō) [you say] for referencing what the other speaker said in their talk more often than the other participants.
觉得 (juéde) 它 (tā) [to think it]	The TESOL participants did not use 觉得 (juéde) [think] with the pronoun 它 (tā) [it] as frequently as the NTESOL participants. The use of 觉得 (juéde) [think] with the pronoun 它 (tā) [it] serves a discourse function mostly.
了解(liǎojiě)它 (tā) [to understand it]	The TESOL participants did not use 它 (tā) as the object when they used the verb 了解 (liǎojiě) [to understand] as often as the NTESOL participants.
它 那 (tā nà) [it that]	The TESOL participants did not use the word 它(tā) [it] with a noun phrase containing the specifier 那 (nà) [that] to refer to something else as often as the LIVT participants.

Table 5.54 Summary of the findings that disagree with Hypothesis 4 (null hypothesis) – Discourse Preference

We can see that the TESOL participants differed in their use of language from the NTESOL participants. The TESOL participants tended to provide more descriptions of time or conditions (时候 [time you]) and tended to discuss what people should do (你 要[you want/must/*BE* going to]) more often than the NTESOL participants. They also referred to what the interviewer said more often than the other groups. (你说[you say]) However, unlike the LIVT participants, the TESOL

participants did not show a strong tendency to use words (adverbs) to add an emphasis, downtoning or hedging in their discourse.

Summary of the findings that Disagree with Hypothesis 4 (null hypothesis) – Lexico-grammatical Differences	
Collocations	**Descriptions OR noteworthy findings/observation**
觉 得 (juéde) 它(tā) [think it]	The TESOL participants did not omit the pronoun 我 (wǒ) [I] in the pattern 我 (wǒ) 觉得 (juéde) 它 (tā) [I think it] as often as the NTESOL participants.
了解 (liǎojiě) 它 (tā) [to understand it]	The TESOL participants did not use 它 (tā) as object when they used the verb 了解 (liǎojiě) [to understand] as much as the NTESOL participants. The TESOL participants did not share similar collocates of 了解 (liǎojiě) 它 (tā) with the LIVT and NTESOL participants.
它 那 (tā nà) [it that]	The TESOL participants did not use the pattern *pronoun + 那 (nà)* [(of) it, that] more often than the LIVT participants. Also, the LIVT and TESOL participants showed a similar frequency in the use of Mandarin possessive marker 的(de) [~'s/of] in the pattern *pronoun + 那(nà) (pronoun + 的*(de) *+ 那(nà))* in which the word 的(de) [~'s/of] is usually omitted. This pattern was used less by the NTESOL participants.

Table 5.55 Summary of the findings that disagree with Hypothesis 4 – Lexico-grammatical Differences

The TESOL participants also used some words/patterns differently from the other participants when observing from a lexico-grammatical perspective. (See Table 5.55.) The TESOL participants, in particular, showed a lower tendency to use 它 [it] as an object (了解 它 [understand it]) in comparison with the other two groups of participants. This finding echos what we discovered about the use of *use it* and *learn it* in chapter 4. The TESOL participants also showed a low use of *it* as an object in the use of these two collocations. This finding regarding the use of 它 [it] suggests the TESOL participants' primings of the use of 它 [it] may be different from the ones of the other participants. The similarity in the use of English *it* and Mandarin 它 [it] in the TESOL group also suggests that the primings of the words could be influenced across languages. In fact, we also found a similar consistency in the use of 我觉得 [I think]*/I think* and 我就 (是) [I just]*/I just* in the LIVT group. Still, this claim of cross-language influences on primings needs more research and evidence to determine its extent and check whether there are other explanations. Nevertheless, what we have found from the discussion above presented is that there is a possibility that these participants had been influenced by some factors so that their primings for certain Mandarin words had become different from the others. In the next chapter, we will investigate the uses of English words *is* and *have* and the Mandarin words 是[to be] and 有[to have].

Chapter 6 Use of IS and 是 (shì) by the three Taiwanese groups

6.1 Introduction

In this chapter, we will investigate the English word *is* and the Mandarin word 是 (shì). The first reason for investigating these words is that they are, not unexpectedly, frequently used words in the Taiwanese groups' English and Mandarin, and the frequently used words provide us with abundant materials for observing how the speakers use them. The word *is* is[5] the 14th most frequent word in the LIVT group, the 11th most frequent word in the TESOL group, and the 12th most frequent word in the NTESOL group. The Mandarin word 是 (shì) is in the top 10 most frequent word lists for all the Taiwanese groups. (See Table 6.1 and Table 6.2.) The second reason is that these words commonly occurr with the person pronouns that we have discussed in chapters 4 and 5, where we saw some noticeable differences between groups, which seemed to warrant further investigation.

IS		N	Freq.	%	Per 1000 words
	LIVT	14	360	1.25	12.47
	TE	11	407	1.37	13.71
	NTE	12	358	1.54	15.36

Table 6.1 Frequencies of *is* in LIVT, TESOL and NTESOL

是 (shì)		N	Freq.	%	Per 1000 words
	LIVT	3	1127	2.92	29.22
	TE	3	941	2.66	26.65
	NTE	4	1139	2.70	27.04

Table 6.2 Frequencies of 是 (shì) in LIVT, TESOL and NTESOL

We combine the use of English *is* and Mandarin 是 (shì) as a set for our investigation. The reason is that, in Chinese-English translation, the English word *is* is often translated to the Mandarin word 是 (shì), because parts of their grammatical description are similar. They are also alike in terms of being used to describe the existence of certain conditions/status. (他 **是 (shì)** 学生 = He **is** a

[5] There are other BE-verbs, such as *are*, *am*, the past tense *was*, *were* and the short form *'s*, *'m* and *'re*. We will see the use of these BE-verbs in the later section. The same applies to the use of the words *has*, *had* and *'ve* in the *have* section.

student.) However, whereas the English *is* (or other BE-verbs) can be used to express tense, aspect, and voice, the Mandarin 是 (shì) does not function that way. The function of the Mandarin 是 (shì) is sometimes to express emphasis, and it often co-occurs with adverbs and then becomes a part of an adverb, such as 就是 (jiùshì) ([*just* + BE] means 'just']. Despite the difference between the English *is* and Mandarin 是 (shì), the fact that they share similar functions makes them a reasonable pair for investigation. Our analysis will of course again focus on finding evidence for or against our hypotheses, and these are repeated here for convenience.

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

6.2 The use of *is*

In the use of *is*, the NTESOL participants showed a tendency to use *is* more in their spoken English. The NTESOL participants used the word *is* 15.36 times per 1000 words, and the LIVT and TESOL participants used the word *is* slightly less often than the NTESOL participants. However, when observing the overall use of auxiliary BE-verbs (*am*, *are*, *is*, *was*, *were* and their short forms), it is found that the three groups did not differ too much in terms of the overall occurrences of auxiliary BE-verbs. There are 1217 occurrences of the auxiliary BE-verbs in total (42.15 times per 1000 words) in the LIVT group. In the TESOL and NTESOL groups, the occurrences of auxiliary BE-verbs are about 44 to 45 times per 1000 words. (See Table 6.3.)

		Freq.	Per 1000 words		LLR	Sig.		
BE-Verbs	LIVT	1217	42.15	LIVT-TE	1.369	0.242		-
	TESOL	1311	44.16	LIVT-NTE	1.943	0.163		-
	NTESOL	1042	44.71	TE-NTE	0.088	0.767		-

Table 6.3 Log-Likelihood Ratios: *BE-verbs* in LIVT, TESOL and NTESOL (1)

Table 6.4 presents the details of how the BE-verbs were used by the participantsof the three groups. We can see that the word *is* and its short form *'s* are the most commonly used *BE* verbs in all the groups. The most obvious differences between the groups are the use of past tense BE-verbs *was* and *were*. The LIVT participants used the word *was* 3.98 times per 1000 words. The TESOL participants used *was* 4.99 times per 1000 words, and the NTESOL participants only used *was* 1.89 times per 1000 times. The gap between these groups is larger in the use of *were*. The LIVT participants used the word *were* 0.69 times per 1000 words, and the TESOL participants used the word *were* 0.20 times per 1000 words. The NTESOL participants only used *were* 0.04 times per 1000 words (only one instance found.) The lack of use of past tense BE-verbs in the NTESOL group matches what we found in the use of the collocation *I was* in chapter 4, namely that the NTESOL participants did not use past tense very often in their spoken English.

	LIVT			TE			NTE		
BE-Verbs	1217	100	42.15	1311	100	44.16	1042	100	44.71
With	Freq.	%	per 1000 words	Freq.	%	per 1000 words	Freq.	%	per 1000 words
am	3	0.25	0.10	14	1.07	0.47	7	0.67	0.30
*'m	54	4.44	1.87	61	4.65	2.05	68	6.53	2.92
are	123	10.11	4.26	156	11.90	5.26	110	10.56	4.72
*'re	8	0.66	0.28	27	2.06	0.91	16	1.54	0.69
is	360	29.58	**12.47**	407	31.05	**13.71**	358	34.36	**15.36**
***'s**	534	43.88	**18.50**	492	37.53	**16.57**	438	42.03	**18.79**
was	115	9.45	**3.98**	148	11.29	**4.99**	44	4.22	**1.89**
were	20	1.64	**0.69**	6	0.46	**0.20**	1	0.10	**0.04**

Table 6.4 Use of aux Be-verbs in LIVT, TESOL and NTESOL

When examining these BE-verbs via log-likelihood ratios, it is found that the LIVT and TESOL participants used the word *was* significantly more often than the NTESOL participants. The LIVT participants used the word *were* significantly more often than the TESOL and NTESOL participants. On the other hand, the LIVT participants used the short form *'re* significantly less often than the TESOL and NTESOL participants. The NTESOL participants used the short form *'m* significantly more often than that LIVT and TESOL participants. (See Table 6.5.) It seems that the NTESOL participants tended to use short forms of auxiliary BE-verbs more often than the LIVT and TESOL participants. This goes against our expectations since the NTESOL participants were the ones who had received the least English input outside their formal English education at school. The use of short forms is usually considered to be associated with casual spoken English. Their strong tendency to use the short forms of the BE-verbs may suggest that the NTESOL participants had been exposed to the use of BE-verbs in short forms '*re* and *'m* frequently. The short forms *'re* and *'m* are often used in the combinations *I'm* and *you're*. If we were able to check whether the EFL learning materials used at school provide frequent instances of *I'm* and *you're*, the significantly more frequent use of the short forms *'re* and *'m* in the NTESOL group might be explainable. Unfortunately, this is beyond the scope of the current research. The other possibility is that it is not the case that the NTESOL participants used the short forms '*m* and *'re* unusually frequently. It may be the LIVT and TESOL participants that used the short forms *'m* and *'re* significantly less often than the NTESOL participants. The LIVT and TESOL participants might be trying to avoid using short forms (*I'm, you're* or *they're*), because they might have wanted to sound more formal in the interview. Either way, the finding suggests that the LIVT and TESOL participants behaved differently from the NTESOL participants.

In the previous paragraph, we mentioned that the LIVT and TESOL participants used the word *was* more often than the NTESOL participants, and the difference is significant. (See Table 6.5.) Since the word *was* is the past tense form of both *am* and *is*, in the next section of investigation of *is* and its short form *'s*, the use of *was* will also be included.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
were	17.243	0.000	***	+	were	8.348	0.004	**	+	**was**	**37.112**	**0.000**	*******	**+**
was	**19.468**	**0.000**	*******	**+**										
Less					Less					Less				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
is	**7.792**	**0.005**	******	**-**	*'re	10.370	0.001	**	-	*'m	3.958	0.047	*	-
*'m	6.004	0.014	*	-	am	7.420	0.006	**	-					
*'re	4.706	0.030	*	-										

Table 6.5 Log-Likelihood Ratios: *BE-verbs* in LIVT, TESOL and NTESOL (2)

6.2.1 L1 Collocates of *is*

In the lists of the top 10 most frequent L1 collocates, we can see that most of the collocates are shared by all the three groups. (See Table 6.6.) The noun *English* and the impersonal pronoun *it* are the most frequent L1 collocates. We can also find many nouns (including gerunds), such as *speaking*, *writing*, *school*, *teacher* and *language*, and the third person pronouns, *he* and *she*, also appear as L1 collocates. The pronouns/nouns/gerunds appear mostly because of the topics in the talks. Therefore, we will not investigate these words further.

We see some differences between the groups with respect to other words in the lists. The LIVT group has the determiners *that* and *this* in their top 10 most frequent L1 collocate list. In the NTESOL group's list, these two determiners are not seen. The TESOL and NTESOL groups have the introductory *there* as a frequent L1 collocate, while it is not seen in the LIVT group's list.

	LIVT	360	12.47	TESOL	407	13.71	NTESOL	358	15.36
N	**L1**	**Freq.**	per 1000 words	**L1**	**Freq.**	per 1000 words	**L1**	**Freq.**	per 1000 words
1	*ENGLISH*	59	2.04	*ENGLISH*	59	1.99	*ENGLISH*	66	2.83
2	IT	26	0.90	IT	33	1.11	IT	27	1.16
3	**THAT**	18	0.62	*HE*	20	0.67	IS	15	0.64
4	**THIS**	13	0.45	IS	19	0.64	*SCHOOL*	14	0.60
5	IS	13	0.45	**THAT**	15	0.51	*HE*	11	0.47
6	*SPEAKING*	9	0.31	*WRITING*	13	0.44	**THERE**	10	0.43
7	*SCHOOL*	9	0.31	EH	11	0.37	*SPEAKING*	10	0.43
8	*REASON*	9	0.31	*TEACHER*	10	0.34	*SHE*	9	0.39
9	*TEACHER*	8	0.28	**THERE**	9	0.30	EH	7	0.30
10	*LANGUAGE*	8	0.28	SPEAKING	8	0.27	*WRITING*	6	0.26

Table 6.6 L1 collocates of *is* in LIVT, TESOL and NTESOL

When checking the use of these L1 collocates via log-likelihood ratios, it is found that the LIVT and TESOL participants used the collocation *that is* significantly more often in comparison with the NTESOL participants. (See Table 6.7.) The LIVT participants used the collocations *this is* and *reason is* significantly more often than the TESOL participants, and used the collocation *he is* significantly less often than the TESOL participants. The TESOL participants used the collocations *school is*, *she is* and *English is* significantly less frequently than the NTESOL participants. As noted previously, the frequent use of *reason*, *school*, *she* and *English* is very likely because of what had been discussed in the interviews. The frequent use of these words with *is* in one group suggests that the participants of this group might tend to talk about these topics more than the other participants. This is probably a difference caused by the different attitudes to English learning that the three groups had in their interviews. Although how EFL learners with different language learning backgrounds view particular topics differently would be a valuable topic for social research, the current study does not concern this aspect. Therefore, we will leave the use of *reason is*, *school is*, *she is* and *English is* untouched. From the remaining collocations *that is* and *this is*, we will select the collocation *that is* for further investigation. The reason is that, as we will see in the following section, the short form of *that is (that's)* is also frequently found in the LIVT and TESOL data. The frequent use of both *that is* and *that's* in these two groups suggests that it may provide us more instances for investigation, and therefore, it potentially provides us with a greater chance to find more evidence with which to test our hypotheses.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
THAT IS	**11.525**	**0.001**	*******	**+**	THIS IS	7.020	0.008	**	+	**THAT IS**	**8.354**	**0.004**	******	**+**
					REASON IS	5.015	0.025	*	+					
					Less					Less				
					Word	LLR	Sig.			Word	LLR	Sig.		
					HE IS	9.224	0.002	**	-	SCHOOL IS	6.894	0.009	**	-
										SHE IS	6.673	0.010	**	-
										ENGLISH IS	3.911	0.048	*	-

Table 6.7 Log-likelihood Ratios: Collocations of *is* – L1 Collocates

When investigating the use of the short form *is ('s)*, it is found that the words *it*, *that, there* and *what* were shared by the three groups. (The words relating to the topic of the talks are marked in italic in Table 6.8 and ignored.) The word *how* only appeared in the LIVT group's list. The word *let* only appeared in the TESOL group's list, and the word *where* only appeared in the NTESOL group's list. (See Table 6.8.)

	LIVT	534	18.50	TESOL	492	16.57	NTESOL	438	18.79
N	**L1**	**Freq.**	**per 1000 words**	**L1**	**Freq.**	**per 1000 words**	**L1**	**Freq.**	**per 1000 words**
1	IT	432	14.96	IT	382	12.87	IT	375	16.09
2	THAT	56	1.94	THAT	59	1.99	THERE	17	0.73
3	THERE	14	0.48	THERE	8	0.27	THAT	15	0.64
4	WHAT	8	0.28	*TEACHER*	6	0.20	*TAIWAN*	7	0.30
5	*TAIWAN*	7	0.24	*HE*	5	0.17	WHAT	3	0.13
6	*SHE*	6	0.21	LET	4	0.13	*SHE*	3	0.13
7	*PEOPLE*	2	0.07	WHAT	3	0.10	*HE*	3	0.13
8	HOW	2	0.07	*TAIWAN*	3	0.10	WHERE	2	0.09
9	*COUNTRY*	2	0.07	*SHE*	3	0.10	*COUNTRY*	2	0.09
10				*STUDENT*	2	0.07			

Table 6.8 L1 collocates of *'s* in LIVT, TESOL and NTESOL

When investigating these data with log-likelihood ratios, it is found that the LIVT and TESOL participants used the combination *that* with the short form *'s* significantly more frequently than the NTESOL participants. The TESOL participants used the combination *it* with the short form *'s* significantly less frequently than the LIVT and NTESOL participants. (See Table 6.9.) We will consider the use of *that's* and *it's* in our further investigation.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
THAT'S	**17.240**	**0.000**	*******	**+**	**IT'S**	**4.623**	**0.032**	*****	**+**	**THAT'S**	**18.409**	**0.000**	*******	**+**
										Less				
										Word	LLR	Sig.		
										IT'S	**9.425**	**0.002**	******	**-**
										THERE'S	5.857	0.016	*	-

Table 6.9 Log-likelihood Ratios: Collocations of *'s* – L1 collocates

With regard to the use of the past tense *was*, it is found that the LIVT and TESOL participants used the collocation *I was* and *it was* significantly more often than the NTESOL participants. In fact, the NTESOL participants used the word *was* significantly less frequently than the LIVT and TESOL participants. Therefore, it is not surprising that the collocations *I was* and *it was* are also infrequently found in the NTESOL data. (See Table 6.10.) We will not investigate the use of *I was* since it has been discussed in chapter 4. We will not investigate the use of *it was*, either, because there is only one instance of *it was* in the NTESOL data. (See Table 6.11.) The low use of *it was* in the NTESOL group is, as we discovered previously in chapter 4 (4.2.2), probably caused by their tendency of not using past tense verbs as often as the LIVT and TESOL participants. In the next sections, we will investigate the use of *it is*/ *it's* and *that is*/*that's*.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
I WAS	12.068	0.001	***	+						I WAS	28.934	0.000	***	+
IT WAS	5.763	0.016	*	+						**IT WAS**	**16.781**	**0.000**	*******	**+**
Less					Less					Less				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
					I WAS	4.287	0.038	*	-					
					IT WAS	**3.978**	**0.046**	*****	**-**					

Table 6.10 Log-likelihood Ratios: Collocations of *was* – L1 collocates

	LIVT	115	3.98	TESOL	148	4.99	NTESOL	44	1.89
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	I	69	2.39	I	98	3.30	I	26	1.12
2	IT	9	0.31	IT	20	0.67	EH	3	0.13
3	THEY	4	0.14	*HE*	6	0.20	*FILMS*	2	0.09
4	THERE	3	0.10	*ENGLISH*	5	0.17	YOU	1	0.04
5	*ENGLISH*	3	0.10	*SHE*	4	0.13	WAS	1	0.04
6	YEAH	2	0.07	THERE	2	0.07	THEY	1	0.04
7	*TIME*	2	0.07	EH	2	0.07	THERE	1	0.04
8	*MAJOR*	2	0.07	WAS	1	0.03	*RIOT*	1	0.04
9	EH	2	0.07	THAT'S	1	0.03	NOT	1	0.04
10	WHAT	1	0.03	THAT	1	0.03	IT	1	0.04

Table 6.11 L1 collocates of *was* in LIVT, TESOL and NTESOL

IT IS / IT'S

With respect to the use of *it is*, it is found that there is no significant difference between the three groups in terms of its frequency. However, with respect to the short form *it's*, it is found that the TESOL participants used *it's* significantly less often than the LIVT and NTESOL participants. (See Table 6.12.) The use of *it's* in the TESOL group is 12.87 times per 1000 words. This is less than the in LIVT group (14.96 times per 1000 words) and in the NTESOL group (16.09 times per 1000 words). (See Table 6.13.)

it is	**LLR**	**Sig.**		
LIVT - TE	0.649	0.420		-
LIVT - NTE	0.839	0.360		-
TE - NTE	0.025	0.874		-
it's	**LLR**	**Sig.**		
LIVT - TE	4.623	**0.032**	*	+
LIVT - NTE	1.057	0.304		-
TE - NTE	9.425	**0.002**	**	-

Table 6.12 Log-likelihood Ratios: *it is* and *it's*

Although a difference exists between the TESOL group and the other two groups in terms of the frequency of *it's*, the way the combination *it's* was used does not differ greatly across groups. In Table 6.13, we can see the patterns of the use of *it is* and *it's*. The patterns show that the three groups share most of the collocates of *it is* and *it's.* The word *think* is the most frequent L1 collocate of both *it's* and *it is*. The connectives *because*, *but*, *and* and *so* are also common L1 collocates for all three groups. The R1 collocates of *it is* and *it's* vary, but the test on the frequencies of these collocates does not reveal any significant difference. This finding suggests that the lack of use of *it's* in the TESOL group may be a different preference in their discourse use of *it's*. The social factor may also be the cause of this difference. The use of *it's* is more colloquial than the use of *it is*. The TESOL participants might be aware of this difference between the use of *it's* and *it is* in speech, and used *it is* for being formal in the interviews. Their English input may be another influence factor. The use of *it is* is relatively strongly associated with writing, while the use of *it's* is almost associated with speech. The English input which the TESOL participants had received mostly is written English. Therefore, they might be influenced to use *it is* more strongly than to use *it's* in their English. This finding, therefore, is potentially compatible with what has been discussed in chapter 2.

When investigating the R1 collocates of *it is* and *it's*, it is found the most frequent R1 collocate of *it is* is the word *because*, and the way these participants used *it is because*/*it's because* shows how the TESOL participants differs from the participants in the LIVT and NTESOL groups.

	LIVT			**TE**			**NTE**		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
IS/'S	894	100	30.97	899	100	30.28	796	100	34.16
IT IS	26	2.91	0.90	33	3.67	1.11	27	3.39	1.16
IT'S	432	48.32	**14.96**	382	42.49	**12.87**	375	47.11	**16.09**
Patterns of *it is*	N L1 Centre R1 1 THINK IT IS BECAUSE 2 IT NOT 3 YES THE 4 BECAUSE EASY 5 SO EH			N L1 Centre R1 1 THINK IT IS HARD 2 EH NOT 3 SO REALLY 4 YES VERY 5 IT EH 6 SAY EASY			N L1 Centre R1 1 THINK IT IS IMPORTANT 2 SO VERY 3 BUT EASY 4 FEEL FOR		
Patterns of *it's*	N L1 Centre R1 1 THINK IT'S NOT 2 EH VERY 3 YEAH EH 4 IT'S THE 5 BECAUSE IT'S 6 MAYBE QUITE 7 BUT EASY 8 IT REALLY 9 SO DIFFICULT 10 AND INTERESTING			N L1 Centre R1 1 THINK IT'S NOT 2 EH VERY 3 IT'S EH 4 AND IT'S 5 THAT REALLY 6 SO USEFUL 7 BUT HARD 8 MAYBE PRETTY 9 BECAUSE THE 10 YEAH OKAY			N L1 Centre R1 1 THINK IT'S EH 2 EH NOT 3 IT'S IT'S 4 BECAUSE VERY 5 SO EASY 6 AND REALLY 7 BUT USEFUL 8 YEAH MORE 9 FEEL IMPORTANT 10 YES HARD		

Table 6.13 Use of *it is* and *it's* in LIVT, TESOL and NTESOL

	LIVT		**TE**		**NTE**	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
IT IS BECAUSE	5	0.17	1	0.03	0	0.00
IT'S BECAUSE	5	0.17	4	0.13	4	0.17
In Total	10	**0.35**	5	**0.17**	4	**0.17**

Table 6.14 Use of *it is because* and *it's because* in LIVT, TESOL and NTESOL

In Table 6.14, it is found that there are 5 occurrences of the collocation *it is because* and 5 occurrences of the collocation *it's because* in the LIVT group. The total use of *it is because* and *it's because* is 10 occurrences (0.35 times per 1000 words). There are not so many occurrences of these

two collocations in the TESOL and NTESOL groups. There are only 1 occurrence of the collocation *it is because* and 4 occurrences of *it's because* in the TESOL group. The total use of *it is because* and *it's because* in TESOL group is 5 occurrences (0.17 times per 1000 words.) The use of *it is because* is not found in the NTESOL group. The collocation *it's because* is found 4 times (0.17 times per 1000 words) in the NTESOL group. From observation of the instances of the collocation *it is because* and *it's because*, it is found that the collocations *it is because*/ *it's because* in the LIVT and NTESOL groups occurr frequently with the words *I think*. (Examples are given in Table 6.16.) Six out of ten instances of *it is because*/*it's because* in the LIVT group occurr with *I think* at the L1 position. Three out of four instances of *it's because* in the NTESOL group occurr as the collocation *I think it's because*. While the LIVT and NTESOL participants showed this tendency to use *I think* with *it is because*/*it's because*, the TESOL participants did not behave in the same way. In all the five instances of *it is because*/*it's because* in the TESOL group, there is no use of *I think* before the collocations. (See Table 6.15.) The log-likelihood ratios of the frequencies of the use of *it is/it's because* does not show any significance to the differences across the groups. In the collocations of *I think it is/it's because*, however, the LIVT participants used this expression significantly more often than the TESOL participants.

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
IT IS/IT'S BECAUSE	10	100	0.35	5	100	0.17	4	100	0.17
I THINK at L1 position	6	60	0.21	0	0	0.00	3	75	0.13

Table 6.15 Use of *I think it is because/it's because* in LIVT, TESOL and NTESOL

What we see in the use of *it is*/*it's* is that the TESOL participants tended not to use the words *it is*/*it's* as often as the LIVT and NTESOL participants did. This different tendency in terms of the frequencies may be caused by the different discourse preference because we cannot find great differences in their use of collocation and lexico-grammatical features. This supports for our hypothesis 2, in which we hypothesised the EFL learner studying in English-relevant subjects will have noticeable differences from those studying in non-English-relevant subjects. In the use of collocates of *it is/it's*, it is found that the use of *I think it is/it's because* was a relatively fixed expression when the LIVT and NTESOL participants used the collocation *it is/it's because*. The finding gives us a chance to see how the LIVT, TESOL and NTESOL participants behaved differently from each other in their use of language.

N	Concordance
1	er people's opinion **I think it is because** learning English i
2	or me yeah . # yeah **I think it is because** when &eh I'm think
3	ned &eh Japanese and Korean **it is because** my also my interes
4	I think it's not so fun but **it is because** I love English so
5	h vocabularies yeah . # yes **it is because** English is first l

Instances of *it is because* from LIVT

N	Concordance
1	t's interesting **I think** yes **it's because** it's different &eh
2	# &eh first of all **I think it's because** the test exam you k
3	think it's easy . # **I think it's because** I I know I'm my my
4	I think it's listening . # **it's because** it &eh English is a
5	p (?) because how can I say **it's because** I have to use Engli

Instances of *it's because* from LIVT

N	Concordance
1	t magazines yeah so **I think it's because** I have a have a goa
2	s . # &eh yes &eh I **I think it's because** I have a interestin
3	h English . # &eh I **I think it's because** I'm I'm very &eh de
4	icult . # I I like them but **it's because** &eh the personal li

Instances of *it's because* from NTESOL

Table 6.16 Instances of the use of *I think it is because/it's because* in LIVT, TESOL and NTESOL

THAT IS / THAT'S

Previously in Table 6.7 and Table 6.9, we saw that the LIVT and TESOL participants used the collocation of *that is* (and the short form *that's*) significantly more frequently than the NTESOL participants. Here we will see how the three groups used the determiner *that* with *is ('s)*.

	LIVT			**TE**			**NTE**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
THAT	278	100	**9.63**	291	100	**9.80**	83	100	**3.56**
THAT IS	18	6.47	0.62	15	5.15	0.51	2	2.41	0.09
THAT 'S	56	20.14	1.94	59	20.27	1.99	15	18.07	0.64
THAT IS/THAT'S	74	26.62	**2.56**	74	25.43	**2.49**	17	20.48	**0.73**
Patterns of *that is*	N L1 Centre R1 1 THINK THAT IS NOT 2 SO INTERESTING 3 IT'S			N L1 Centre R1 1 THINK THAT IS VERY 2 AND THE			*[Inapplicable]*		
Patterns of *that's*	N L1 Centre R1 1 SO THAT'S ALL 2 YEAH THE 3 LANGUAGE WHY 4 EH REALLY 5 THAT NOT 6 THINK IT 7 THAT'S SO 8 ENGLISH THAT'S 9 IT 10 MAYBE			N L1 Centre R1 1 THINK THAT'S THE 2 YEAH IT 3 SO VERY 4 EH INTERESTING 5 AND PRETTY 6 UNDERSTAND IT'S 7 BECAUSE WHY 8 DIFFERENT			N L1 Centre R1 1 AND THAT'S NOT 2 EH EH 3 BECAUSE ALL		

Table 6.17 Use of *that is* and *that's* in LIVT, TESOL and NTESOL

In Table 6.17, we can see that the LIVT and TESOL participants used the words *that, that is* and *that's* similarly in terms of their frequencies. The determiner *that* with *is* ('s) occurrs 74 times in the LIVT data (2.56 times per 1000 words.) It also occurrs 74 times in the TESOL data (2.49 times per 1000 words.) There are only 17 occurrences of *that* with *is ('s)* in the NTESOL data (0.73 times per 1000 words.) The LIVT and TESOL participants used the collocation *that is ('s)* twice as frequently as the NTESOL participants did. In fact, the LIVT and TESOL participants also used the word *that* twice as often as the NTESOL participants. The use of *that* in the LIVT group is 9.63 times per 1000 words. The use of *that* in the TESOL group is 9.80 times per 1000 words. The use of *that* in the NTESOL is 3.56 times per 1000 words. When taking a closer look at the patterns in which the collocations *that is* and *that's* occur, it is found that the LIVT participants and the TESOL participants

are alike. There were several collocates shared within the LIVT and TESOL groups, such as the verb *think*, the connective *so*, the article *the* and the *wh*-adverb *why*. These collocates were not frequently used by the NTESOL participants. The finding that the LIVT and TESOL participants differ from the NTESOL participants in the use of *that* and *that is('s)* partly supports the two hypotheses that the EFL learners who go to the UK for study and those who study in English-relevant subjects will have noticeable differences in their English use in comparison to those who neither go abroad for study nor major in English-relevant subjects.

	BASEah			**COCAsp**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
THAT	8253	100	19.53	354960	100	19.24
THAT IS	260	3.15	0.62	11041	3.11	0.60
THAT 'S	1149	13.92	2.72	44078	12.42	2.39
Patterns of *that is*	N L1 Centre R1 1 AND THAT IS THE 2 THAT THAT 3 IS TO 4 SO WHY 5 SOMETHING AN 6 THINK NOT 7 BUT ONE 8 ALL WHAT 9 WHATEVER REALLY 10 DO SOMETHING			1 AND THAT IS THE 2 THINK NOT 3 THAT WHAT 4 OF GOING 5 SOMETHING THAT 6 BUT SO 7 SO TO 8 KOTB IT 9 IS JUST 10 WELL N'T		
Patterns of *that's*	N L1 Centre R1 1 AND THAT'S THE 2 SO WHAT 3 THINK NOT 4 THAT RIGHT 5 BUT ONE 6 THAT'S WHY 7 YEAH THAT'S 8 KNOW WHERE 9 OKAY JUST 10 BECAUSE VERY			N L1 Centre R1 1 AND THAT 'S WHAT 2 THINK THE 3 BUT NOT 4 SO RIGHT 5 WELL WHY 6 THAT GOING 7 KNOW WHERE 8 BECAUSE ALL 9 KOTB HOW 10 CAMERA IT		

Table 6.18 Use of *that is* and *that's* in BASEah and COCAsp

When comparing the use of *that is('s)* by the LIVT and TESOL groups to those shown in the reference corpora, it is found that the LIVT and TESOL groups are similar to the BASEah and COCAsp corpora in terms of their collocates. The verb *think*, the connective *so*, the article *the* and the *wh*-adverb *why* often co-occur with the collocation *that is* and *that's*. (See 6.18.) It seems that the use of *that is* and *that's* does not differ greatly in the native corpora, and that the LIVT and TESOL participants used the words *that's* and *that is* in a way similar to those found in the native corpora.

The findings relating to *that is* and *that's* are therefore that (1) the LIVT and TESOL participants used the word *that* and the collocation *that is ('s)* more often than the NTESOL participants, and (2) their collocates of *that is ('s)* were similar to the ones found in the reference corpora. The findings support for hypothesis1 and 2. The explanation for these differences between the LIVT and TESOL participants and the NTESOL participants (and the similarity between the LIVT and TESOL participants and the reference corpora) is very likely to relate to the different English input to which the participants of these groups had been exposed. The word *that* is a common word in both academic spoken English (BASEah) and semi-formal (casual) spoken English (COCAsp), and is also commonly seen in written English. The LIVT and TESOL participants, who had had plenty of opportunities for exposure to spoken English and written English and for using both in either their studies or daily lives, are likely to have been influenced by this exposure to use the collocation *that is ('s)* in a manner similar to that of native speakers. The NTESOL participants, on the other hand, who had had fewer opportunities to receive and produce English in their studies and daily lives, are likely not to had been influenced to use the word *that* and *that is* ('s) in the same way as the LIVT and TESOL participants. This finding is, to some degree, compatible with what has been discussed in chapter 2. Comprehensible input, meaningful output along with the practice of negociation are the key elements which influence people's language learning. When the use of words occurs frequently, it is likely to be stored in a person's long-term memory and easier to be retrieved afterwards. Therefore, people are likely to be strongly primed to use the words/combination of words to which they are frequently exposed. The LIVT and TESOL participants had had greater opportunities to be exposed to and practice to this kind of input/output, so it is very likely their language use was influenced accordingly. What we have observed suggests that the immersion in different language environments and the diversity of school education are very likely to cause people to be primed differently in their language use.

6.2.2 R1 Collocates of *is*

We now turn to investigate the R1 collocates of *is*. Table 6.19 presents the top 10 most frequent R1 collocates of *is*, and Table 6.20 presents the top 10 most frequent R1 collocates of the short form *'s*. There are several R1 collocates shared by all the three groups. The negative marker *not*, the articles *a* and *the*, and the adverbs *very* and *really* are frequently found in all the three groups. There are several words, though, that only occurr in one or two groups' top 10 frequent R1 collocate lists. For instance, the adjective *easy* appears in the LIVT and NTESOL groups' top 10 most frequent R1 collocate lists but not in the TESOL list. The word *quite* only appears in the LIVT group's list. Likewise, the word *that* appears only in the TESOL group's list. The finding that some collocates appear in (or are missing from) one group's top 10 frequent collocate list suggests that there are potentially some noticeable differences between groups for testing our hypotheses 1 and 2.

In the lists, several adjectives, such as *easy*, *important* and *useful*, occur in high frequency. The frequent use of these words is probably a result of the topics that were discussed in the interviews. The participants were asked about which parts of language learning were easy or difficult for them, and how important English would be for their career. Therefore, we will not investigate this kind of use any further.

	LIVT	360	12.47	TESOL	407	13.71	NTESOL	358	15.36
N	**R1**	**Freq.**	**per 1000 words**	**R1**	**Freq.**	**per 1000 words**	**R1**	**Freq.**	**per 1000 words**
1	THE	33	1.14	NOT	44	1.48	EH	39	1.67
2	A	28	0.97	EH	36	1.21	VERY	35	1.50
3	NOT	26	0.90	VERY	35	1.18	NOT	31	1.33
4	EH	19	0.66	A	35	1.18	A	28	1.20
5	VERY	16	0.55	THE	23	0.77	THE	17	0.73
6	IS	13	0.45	IS	19	0.64	IS	15	0.64
7	BECAUSE	11	0.38	THAT	13	0.44	*USEFUL*	10	0.43
8	QUITE	10	0.35	MORE	12	0.40	REALLY	10	0.43
9	FROM	8	0.28	REALLY	11	0.37	*IMPORTANT*	10	0.43
10	*EASY*	8	0.28	*INTERESTING*	11	0.37	*EASY*	10	0.43

Table 6.19 R1 collocates of *is* in LIVT, TESOL and NTESOL

	LIVT	534	18.50	TESOL	492	16.57	NTESOL	438	18.79
N	**R1**	**Freq.**	**per 1000 words**	**R1**	**Freq.**	**per 1000 words**	**R1**	**Freq.**	**per 1000 words**
1	NOT	41	1.42	A	47	1.58	EH	43	1.85
2	VERY	31	1.07	NOT	46	1.55	A	42	1.80
3	EH	27	0.94	VERY	29	0.98	NOT	35	1.50
4	A	26	0.90	EH	25	0.84	IT'S	29	1.24
5	THE	22	0.76	IT'S	18	0.61	VERY	24	1.03
6	IT'S	22	0.76	REALLY	15	0.51	*EASY*	12	0.51
7	QUITE	16	0.55	*USEFUL*	14	0.47	REALLY	11	0.47
8	REALLY	14	0.48	PRETTY	13	0.44	*USEFUL*	10	0.43
9	*EASY*	14	0.48	HARD	13	0.44	MORE	10	0.43
10	ALL	13	0.45	THE	10	0.34	*IMPORTANT*	10	0.43

Table 6.20 R1 collocates of *'s* in LIVT, TESOL and NTESOL

In Table 6.21 and Table 6.22, it is found that the LIVT participants used the word *quite* as a R1 collocate with *is ('s)* significantly more often than the TESOL and NTESOL participants. The TESOL participants on the other hand used the word *pretty* with *is ('s)* significantly more often than the LIVT and NTESOL participants. It is intriguing that the participants of one group preferred to use some adverbs with *is ('s)* significantly more often than the ones of the other groups[6]. Therefore, we will take a closer look on how these participants used *is ('s) quite* and *is ('s) pretty* in the later section.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
IS QUITE	**4.246**	**0.039**	*****	**+**	IS BECAUSE	7.112	0.008	**	+	*IS INTERESTING*	7.507	0.006	**	+
					IS QUITE	**6.047**	**0.014**	*****	**+**	**IS THAT**	**6.572**	**0.010362**	*****	**+**
					IS FROM	4.024	0.045	*	+					
Less					**Less**					**Less**				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
IS EH	11.986	0.001	***	-	**IS VERY**	**6.733**	**0.009**	******	-	*IS USEFUL*	7.933	0.005	**	-
IS VERY	**11.904**	**0.001**	*******	-	IS EH	4.879	0.027	*	-	*IS EASY*	4.313	0.038	*	-
IS USEFUL	5.624	0.018	*	-	IS NOT	4.193	0.041	*	-					

Table 6.21 Log-likelihood Ratios: Collocations of *is* – R1 Collocates

[6] The TESOL and NTESOL participants used the word *very* with *is* significantly more frequently than the LIVT participants. However, when the use of *'s very* is taken into calculation, the significance disappears. Therefore, we will not include the use of *is('s) very* in our investigation.

LIVT - NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
'S QUITE	**9.603**	**0.002**	******	**+**	**'S QUITE**	**8.048**	**0.005**	******	**+**	**'S PRETTY**	**9.504**	**0.002**	******	**+**
'S THE	4.139	0.042	*	+	'S THE	4.952	0.026	*	+					
Less					Less					Less				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
'S A	8.004	0.005	**	-	*'S HARD*	11.872	0.001	***	-	'S EH	10.170	0.001	**	-
'S EH	7.917	0.005	**	-	**'S PRETTY**	**8.711**	**0.003**	******	**-**	'S MORE	7.933	0.005	**	-
					'S A	5.557	0.018	*	-	'S IT'S	5.947	0.015	*	-
					'S USEFUL	4.192	0.041	*	-					

Table 6.22 Log-likelihood Ratios: Collocations of *'s* – R1 Collocates

The LIVT participants used the article *the* with *is ('s)* significantly more frequently than the TESOL and NTESOL participants did, but used the article *a* with *is ('s)* significantly less frequently than the TESOL and NTESOL participants. With respect to the use of the past tense *was*, the TESOL participants used the article *a* significantly more often than the LIVT and NTESOL participants. (See Table 6.23.) When checking the instances of the use of *is ('s) a* and *is ('s) the*, it is found that the words *a* and *the* were used in a similar way. The word *a* was often used in the collocation *is ('s) a little/a lot*, and the word *the* was used in the collocations *is ('s) the most*/*the same* shared by the three groups. The words *a* and *the* were also used in their basic function *a/the + noun*. The findings reveal little lexico-grammatical difference in the use of *is ('s) a/the* between groups. Therefore, we will not present the use of *is ('s) a*/*the* in detail due to limitations of space.

LIVT – NTE					**LIVT - TE**					**TE - NTE**				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
WAS IN	5.923	0.015	*	+						WAS IN	12.184	0.000	***	+
WAS NOT	3.869	0.049	*	+						WAS A	11.419	0.001	***	+
Less					Less					Less				
Words	LLR	Sig.			Word	LLR	Sig.			Words	LLR	Sig.		
					WAS A	5.557	0.018	*	-					

Table 6.23 Log-likelihood Ratios: Collocations of *was* – R1 Collocates

It is also found that there are noticeable differences in the use of the collocations *is/'s eh*, *is because, is from, is that*, *is not, is/'s useful, is easy, 's hard, 's more* and *'s it's* in terms of their frequencies. As we have mentioned, the frequent use of the words *easy*, *useful* and *hard* is very likely related to the contents of our interviews, so we will not further investigate their use. The use of repetition ([*it*]*'s it's*), the filler sound (*is/'s eh*) and negation (*is not*) will likewise not be investigated in the current study. We will also not include *is because* in our further investigation, because the words *is because* commonly occurs as part of the combination *it is because*, and we already have undertaken an investigation of the use of *it's because* in the previous section.

The LIVT participants used *is from* significantly more often than the TESOL participants. However, when checking how this collocation was used, it is found that there is no great difference in their lexico-grammatical use. The frequent use of *is from* in the LIVT data shows that the LIVT participants tended to talk about the place/person where an item was originally located or by whom it was originally possessed more often than the TESOL participants. Since it is a difference of discourse preference, we will not go into depth about its use. A similar situation applies to the use of *'s more*. The NTESOL participants used this collocation more often than the TESOL participants, and the examination of the instances suggests that the NTESOL participants tended to make comparisons more often than the TESOL participants. This is also a difference of discourse preference, so we will not go into depth about its use. The TESOL participants used the collocation *is that* significantly more often than the NTESOL participants. Previously in chapter 4.2.2, we discovered that the TESOL participants had a tendency to use *think that* more often than the other participants. The word *that* here is in its linking function of connecting the subordinate clause to the main clause. The frequent use of *is that* in the TESOL group may be the same situation. We will investigate how the collocation *is that* is used by the three groups in the following section.

IS THAT

The TESOL participants used the collocation *is that* significantly more often than the NTESOL participants. The use of *is that* is found 13 times (0.44 times per 1000 words) in the TESOL group. The use of *is that* is found 5 times (0.17 times per 1000 words) in the LIVT group, and twice (0.09 times per 1000 words) in the the NTESOL group. The TESOL participants used *is that* almost 9 times as often as the NTESOL participants, and over twice as often as the LIVT participants. (See Table 6.25.) When observing the instances of *is that* used by the three groups, it is found that there are three kinds of use. The first kind of use is connecting a noun phrase with a clause. (See Example 1 in Table

6.24.) The second kind is the interrogative use of *is that*. (See Example 2 in Table 6.24.) The third kind is the repetition of the words *that BE*. (See Example 3 in Table 6.24.)

1	I think the most serious problem **is that** &eh (...) &eh they
2	it's multiple question and **is that** a strategy? # and
3	ve us some activities that **is that** was more easy much

Table 6.24 Examples of the use of *is that* in LIVT, TESOL and NTESOL

In Table 6.25, we can see that for the second and the third kinds of use of *is that*, there are no great differences in frequency. It is the first kind of use of *is that* shows how different the TESOL participants were from the participants of the other two groups. 84.62% of the total use of *is that* in the TESOL group is of the first kind of use - *NP+ is that + [S]* (0.37 times per 1000 words.) The LIVT participants used the first kind of use 0.10 times per 1000 words, and the NTESOL participants only used the first kind of use 0.04 times per 1000 words. The TESOL participants used the pattern *NP+ is that + [S]* more frequently than the NTESOL and LIVT participants. Checking the log-likelihood ratios of this use across groups, it is found that the TESOL participants used the pattern *NP+ is that + [S]* significantly more often than the LIVT and NTESOL participants. (See Table 6.26.)

		LIVT			TE			NTE		
		Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
	is that	5	100	0.17	13	100	0.44	2	100	0.09
1	NP + *is that* + [S]	3	**60**	**0.10**	11	**84.62**	**0.37**	1	**50**	**0.04**
2	*(what) is that/ is that*?	2	40	0.07	1	7.69	0.03	1	50	0.04
3	Repetition *(that is that)*	0	0	0	1	7.69	0.03	0	0	0

Table 6.25 Use of *is that* in LIVT, TESOL and NTESOL

NP+ is that + [S]	**LLR**	**Sig.**		
LIVT - TE	4.64	**0.031**	*	-
LIVT - NTE	0.66	0.415		+
TE - NTE	7.51	**0.006**	**	+

Table 6.26 Log-Likelihood Ratios: *NP+ is that + [S]*

Our investigation has shown that the TESOL participants used *is that* in the pattern NP+ *is that* + [S] significantly more often than the LIVT and NTESOL participants. This finding supports for our hypothesis 2, in which we hypothesised that the EFL learners studying in English-relevant subjects will have noticeable differences from the ones studying in non-English-relevant subjects. It can be considered a lexico-grammatical difference. The TESOL participants were primed strongly to use *is* and *that* for the use of this pattern, and the LIVT and NTESOL participants (especially the NTESOL participants) were not primed as much. The explanation for this great difference is likely to be relevant to the English input to which these TESOL participants had been exposed. The TESOL participants, as we have mentioned early, had the academic and written style English as their primary input, while the LIVT participants had received both academic/casual and spoken/written English at the same time. The NTESOL participants had the fewest opportunities to receive such English input. It is possible that the academic and written style of English input had influenced on how the TESOL participants used their English. Therefore, it might be that the TESOL participants' priming of the use of *that* differed from the others' priming. The findings also match what we found with regard to the use of *think that*, where the TESOL participants were also found to use *that* differently from the other participants. The findings are compatible with what we has discussed in chapter 2. Comprehensible input can arise during formal classroom teaching or in informal daily conversation. The "formal and informal environment contribute to second language competence in different ways or rather, to different aspects of second language competence" (Krashen, 1988: 47). Here we can see the potential evidence for the fact that different aspects of language competence can be influenced by how learners are trained in class. Of course, if we want to have a firm statement that the influence is directly a result of their input, a thorough collection and examination of their English input would be needed. Nevertheless, what we found here does support our hypothesis 2, and it also shows a possibility that the input may be the factor behind the differences.

IS QUITE/'S QUITE

The LIVT participants tended to use the adverb *quite* with *is* (and the short form *'s*) significantly more frequently than the TESOL and NTESOL participants. (See Table 6.27.) There are 27 occurrences of *is quite/'s quite* found in the LIVT data (0.94 times per 1000 words). There are only 7 occurrences of the use of *is quite/'s quite* found in the TESOL data (0.24 times per 1000 words) and 5 occurrences of the use of *is quite/'s quite* in the NTESOL data (0.21 times per 1000 words.) The LIVT participants used the collocation *is quite/'s quite* almost 4 times more often than the TESOL and NTESOL participants. In fact, the use of the adverb *quite* in general is also significantly more

frequent in the LIVT group than in the TESOL and NTESOL groups. There are 44 occurrences of the adverb *quite* found in the LIVT data (1.52 times per 1000 words.) On the other hand, there are 17 occurrences of the adverb *quite* in the TESOL data (0.57 times per 1000 words) and 12 occurrences of the adverb *quite* in the NTESOL data (0.51 times per 1000 words.) The use of adverb *quite* in the LIVT group is roughly three times more frequent than the use of *quite* in the TESOL and NTESOL groups. (See Table 6.27.)

is ('s) quite	LLR	Sig.		
LIVT - TE	13.12	**0.000**	***	+
LIVT - NTE	12.28	**0.000**	***	+
TE - NTE	0.14	0.710		+
quite	LLR	Sig.		
LIVT - TE	13.14	**0.000**	***	+
LIVT - NTE	13.23	**0.000**	***	+
TE – NTE	0.08	0.777		+

Table 6.27 Log-Likelihood Ratios: *quite* and *is/'s quite*

	LIVT			**TESOL**			**NTESOL**		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
QUITE	44	100	**1.52**	17	100	**0.57**	12	100	**0.51**
IS QUITE	10	22.73	0.35	2	11.76	0.07	2	16.67	0.09
'S QUITE	17	38.64	0.59	5	29.41	0.17	3	25.00	0.13
IS/'S QUITE	27	61.36	**0.94**	7	41.18	**0.24**	5	41.67	**0.21**
Patterns of *is quite*	N L1 Centre R1 1 SPEAKING IS QUITE GOOD			N L1 Centre R1 1 HE IS QUITE					
Patterns of *'s quite*	N L1 Centre R1 1 THINK IT'S QUITE USEFUL 2 SO GOOD 3 YEAH EH 4 BECAUSE			N L1 Centre R1 1 THINK IT'S QUITE USEFUL			N L1 Centre R1 1 THINK IT'S QUITE		

Table 6.28 Use of *quite* and *is quite* in LIVT, TESOL and NTESOL

Although the LIVT participants used the collocation *is ('s) quite* significantly more often than the other participants, the way the LIVT participants used the collocation *is ('s) quite* did not differ too greatly from the way the TESOL and NTESOL participants did. When investigating on the collocates with the collocation *is ('s) quite*, it is found that it shows similar patterns in all the groups. The collocations *is ('s) quite good* and *is ('s) quite useful* were used frequently by the participants of the three groups, even though these collocations differed in the degree of frequency. (See Table 6.28.)

In fact, the adjectives following the collocation *is ('s) quite* are found to be similar in all the three groups. Examining the data, it is found that the adverb *quite* in the instances of *is ('s) quite* is used to modify positive adjectives mostly. A large proportion of the adjectives contains positive meaning. The exceptions were that there is one occurrence of the neutral adjective *different* in the LIVT and NTESOL data and there is one occurrence of the negative adjective *horrible* in the LIVT data. (See Table 6.29.) In the reference corpora, the adverb *quite* in the instances of *is ('s) quite* is also used to modify positive adjectives. However, we found that the adverb *quite* is also used to modify many neutral adjectives in the reference corpora. (See Table 6.30.)

is / 's quite	LIVT	TESOL	NTESOL
Positive adj.	good	useful	useful
	useful	open-minded	easy
	interesting	important	good
	helpful	true	fresh
	easy		
	important		
Neutral adj.	different	different	
Negative adj.	horrible		

Table 6.29 Use of adjectives with *is / 's quite* in LIVT, TESOL and NTESOL

is / 's quite	BASEah	COCAsp[7]
Positive adj.	interesting	right
	important	remarkable
	right	good
	clear	important
		interesting
		clear
		correct
		amazing
		true
		significant
Neutral adj.	ambiguous	extraordinary
	different	different
	typical	possible
	large	unusual
	subtle	strong
	complicated	small
		real
		high
Negative adj.	difficult	difficult

Table 6.30 Use of adjectives with *is ('s) quite* in BASEah and COCAsp

[7] Because there are many instances of *is('s) quite* in the COCAsp corpus (499 occurrences), we ignored adjectives that occur less than 3 times.

These findings suggest that even though the LIVT participants used the collocation *is ('s) quite* significantly more frequently than the TESOL and NTESOL participants, the way they used this collocation did not differ very much. However, the fact that the LIVT participants used the adverb *quite* to modify adjectives significantly more often than the TESOL and NTESOL participants provides us some support for our hypothesis 1. The similarity of the collocation/colligation (lexico-grammatical use)/semantic associations in the use of *is ('s) quite/quite* suggests that this is probably a difference of discourse preference. We can find examples of *is ('s) quite/quite* in the three groups. However, the different strength in the use of *quite* for modification suggests that it is possible that some factors had influenced the LIVT participants to behave differently from the TESOL and NTESOL participants. It is also very possible that, since these groups are the collection of individuals with similar backgrounds, the explanation of the significant difference may not simply be the weakness or strength of one's priming for using certain words which one possesses, but a case of more or fewer primed individuals in one group as opposed to another groups. Either way, what we found in the use of *is ('s) quite* in the LIVT group shows that a difference exists between the EFL learners with different English learning backgrounds in their language use. We then look at the use of *is pretty/'s pretty* in the three groups.

IS PRETTY/'S PRETTY

In Table 6.31, we can see that the TESOL participants used the collocation *is ('s) pretty* significantly more frequently than the LIVT and NTESOL participants[8]. In fact, the TESOL participants also used the word *pretty* significantly more often than the LIVT and NTESOL participants. There are 23 occurrences of *is ('s) pretty* found in the TESOL data (0.77 times per 1000 words.) There are only 4 occurrences of *is ('s) pretty* in the LIVT data (0.14 times per 1000 words) and none at all in the NTESOL data. The use of *is ('s) pretty* in the TESOL group is 5 times more frequent than in the LIVT group. There are 41 occurrences of the word *pretty* found in the TESOL data (1.38 times per 1000 words.) There are 9 occurrences of *pretty* found in the LIVT data (0.31

[8] There is no instance of the word *pretty* found in the NTESOL group. However, with the number 0, it is impossible to apply the log-likelihood ratios to check the significance, so we insert a virtual 1 occurrence in the NTESOL group for calculation. If it shows significance, then we assume that the zero occurrence of this use in the NTESOL group also shows significance.

times per 1000 words) and again none in the NTESOL data. The use of the word *pretty* in the TESOL group is over 4 times more frequent than in the LIVT group. (See Table 6.32.)

is / 's pretty	LLR	Sig.		
LIVT – TE	14.25	**0.000**	***	-
LIVT – NTE	1.34	0.247		+
TE – NTE	19.98	**0.000**	***	+
pretty	LLR	Sig.		
LIVT – TE	21.29	**0.000**	***	-
LIVT – NTE	5.76	**0.016**	*	+
TE – NTE	39.71	**0.000**	***	+

Table 6.31 Log-Likelihood Ratios: *pretty* and *is / 's pretty*

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
PRETTY	9	100	**0.31**	41	100	**1.38**	0	0	0
IS PRETTY	2	22.22	0.07	7	17.07	0.24	0	0	0
'S PRETTY	2	22.22	0.07	16	39.02	0.54	0	0	0
IS/'S PRETTY	4	44.44	**0.14**	23	56.10	**0.77**	0	0	0
Patterns of *is pretty*				N L1 Centre R1 1 ENGLISH IS PRETTY					
Patterns of *'s pretty*				N L1 Centre R1 1 THINK IT'S PRETTY USEFUL 2 THAT'S PRETTY HARD 3 FUN					

Table 6.32 Use of *pretty* and *is ('s) pretty* in LIVT, TESOL and NTESOL

When checking the use of *pretty* and the collocation *is ('s) pretty* in the reference corpora, it is found that the use of *pretty* is 0.16 times per 1000 words in the BASEah corpus, and 0.33 times per 1000 words in the COCAsp corpus. The use of *is ('s) pretty* is even less frequent. There are only 0.03 instances per 1000 words of the use of *is ('s) pretty* in the BASEah corpus, and 0.06 instances per 1000 words of the use of *is ('s) pretty* in the COCAsp corpus. It seems that in academic spoken English (BASEah) and semi-formal/casual spoken English (COCAsp), the word *pretty* and the collocation *is ('s) pretty* are not commonly used. (See Table 6.33.)

<table>
<tr><th></th><th colspan="3">BASEah</th><th colspan="3">COCAsp</th></tr>
<tr><th></th><th>Freq.</th><th>%</th><th>Per 1000 words</th><th>Freq.</th><th>%</th><th>Per 1000 words</th></tr>
<tr><td>PRETTY</td><td>68</td><td>100</td><td>0.16</td><td>6141</td><td>100</td><td>0.33</td></tr>
<tr><td>IS PRETTY</td><td>6</td><td>8.82</td><td>0.01</td><td>372</td><td>6.06</td><td>0.02</td></tr>
<tr><td>‘S PRETTY</td><td>5</td><td>7.35</td><td>0.01</td><td>652</td><td>10.62</td><td>0.04</td></tr>
<tr><td>IS/’S PRETTY</td><td>11</td><td>16.18</td><td>0.03</td><td>1024</td><td>16.67</td><td>0.06</td></tr>
<tr><td>Patterns of is pretty</td><td colspan="3"><table><tr><th>N</th><th>L1</th><th>Centre</th><th>R1</th></tr><tr><td>1</td><td>WHICH</td><td>IS PRETTY</td><td>MUCH</td></tr><tr><td>2</td><td>THIS</td><td></td><td></td></tr></table></td><td colspan="3"><table><tr><th>N</th><th>L1</th><th>Centre</th><th>R1</th></tr><tr><td>1</td><td>THIS</td><td>IS PRETTY</td><td>MUCH</td></tr><tr><td>2</td><td>WHICH</td><td></td><td>GOOD</td></tr><tr><td>3</td><td>IT</td><td></td><td>AMAZING</td></tr><tr><td>4</td><td>THAT</td><td></td><td>CLEAR</td></tr><tr><td>5</td><td>THINK</td><td></td><td>COOL</td></tr><tr><td>6</td><td>NOW</td><td></td><td>FUNNY</td></tr><tr><td>7</td><td>ONE</td><td></td><td>SIMPLE</td></tr><tr><td>8</td><td>EVERYBODY</td><td></td><td>HARD</td></tr><tr><td>9</td><td>EVERYTHING</td><td></td><td>REMARKABLE</td></tr><tr><td>10</td><td>THEM</td><td></td><td>BAD</td></tr></table></td></tr>
<tr><td>Patterns of ‘s pretty</td><td colspan="3"><table><tr><th>N</th><th>L1</th><th>Centre</th><th>R1</th></tr><tr><td>1</td><td></td><td>THAT'S PRETTY</td><td>GOOD</td></tr><tr><td>2</td><td></td><td>IT'S PRETTY</td><td></td></tr></table></td><td colspan="3"><table><tr><th>N</th><th>L1</th><th>Centre</th><th>R1</th></tr><tr><td>1</td><td>IT</td><td>S PRETTY</td><td>MUCH</td></tr><tr><td>2</td><td>THAT</td><td></td><td>CLEAR</td></tr><tr><td>3</td><td>SHE</td><td></td><td>GOOD</td></tr><tr><td>4</td><td>HE</td><td></td><td>COOL</td></tr><tr><td>5</td><td>THERE</td><td></td><td>AMAZING</td></tr><tr><td>6</td><td>WHAT</td><td></td><td>OBVIOUS</td></tr><tr><td>7</td><td>YOU</td><td></td><td>HARD</td></tr><tr><td>8</td><td>ANNE</td><td></td><td>FUNNY</td></tr><tr><td>9</td><td>SMUCKER</td><td></td><td>BAD</td></tr><tr><td>10</td><td></td><td></td><td>CLOSE</td></tr></table></td></tr>
</table>

Table 6.33 Use of *pretty* andis (‘s) *pretty* in BASEah and COCAsp

When observing how the collocation *is (‘s) pretty* was used, it is found that the patterns in the TESOL data are not very similar to the ones in the LIVT data or the reference corpora. There are only two adjectives *hard* and *different* shared by the LIVT and TESOL groups. The most frequent collocate of the collocation *is (‘s) pretty* in the reference corpora is the word *much*. However, there is no occurrence of *is (‘s) pretty much* found in either the TESOL group or the LIVT group. In fact, when checking the adjectives modified by *pretty* in the instances, it is hard to see too much in common. The similarity we found is that the TESOL participants and the speakers in the reference corpora tended to use the word *pretty* to modify the positive adjectives. (See Table 6.34.) The findings of the use of *is(‘s) quite* and *is(‘s) pretty* are compatible with what we have discussed in chapter 2. It is an example of ‘semantic prosody’ (Sinclair, 2004). The term ‘semantic prosody’ means that certain words can be perceived with positive or negative association via frequent occurrences with specific collocations. In this case, the collocations *is(‘s) quite* and *is (‘s) pretty* were perceived with positive

association. Although it is hard to state for sure that the input to which the TESOL participants had received directly caused this difference, what we found here shows a possibility that the input may be the factor behind the differences. To prove this statement, a thorough collection and examination of their English input would be needed.

is / 's pretty		Freq.		Freq.		Freq.
LIVT	**Positive adj.**	**1**	**Neutral adj.**	**3**	**Negative adj.**	**0**
	well	1	young	1		
			hard	1		
			different	1		
TESOL	**Positive adj.**	**10**	**Neutral adj.**	**6**	**Negative adj.**	**4**
	useful	3	hard	3	boring	1
	fun	3	different	2	poor	1
	important	2	professional	1	sad	1
	interesting	1			bad	1
	nice	1				
BASEah	**Positive adj.**	**5**	**Neutral adj.**	**4**	**Negative adj.**	**1**
	good	3	much	2	absurd	1
	important	1	impenetrable	1		
	clear	1	dark	1		
COCAsp[9]	**Positive adj.**	**267**	**Neutral adj.**	**216**	**Negative adj.**	**28**
	clear	68	much	146	bad	16
	good	61	hard	19	scary	7
	cool	43	obvious	15	frightening	5
	amazing	36	simple	15		
	funny	19	close	11		
	significant	10	unusual	5		
	easy	8	strong	5		
	impressive	7				
	great	5				
	well	5				
	remarkable	5				

Table 6.34 Use of adjectives with *is ('s) pretty* in LIVT, TESOL, BASEah and COCAsp

[9] Because there are too many instances of *is('s) pretty* in the COCAsp corpus (1024 occurrences), we neglected the adjectives that occur less than 5 times.

Similar to what we have found concerning the use of *is ('s) quite*, the finding that the TESOL participants used the collocation *is ('s) pretty/pretty* significantly more frequently than the LIVT and NTESOL participants supports our hypothesis 2 that the EFL participants studying in English relevant subjects will behave differently from those studying in non-English-relevant subjects. The word *pretty* was used as a discourse means to modify the degree of the adjectives. The frequent use of *pretty* for modification in the TESOL group's spoken English suggests that there are some factors influencing them which the LIVT and NTESOL participants did not share with the TESOL participants. In this case, the TESOL participants' different English input could be the possible explanation. In our discussion of the use of *is ('s) quite/quite*, we commented that a significant difference may not simply be the weakness or strength of priming for using certain words that a person possesses, but possibly a case of more or fewer primed individuals in one group using the word(s) as opposed to another groups. With this in mind we checked the data and found that participants LIVT03 and TE01 contributed almost 50% of the total use of the collocation *is ('s) quite* and *is ('s) pretty*. It is possible that these two participants were primed strongly for such usages in their spoken English. Although we have not got access to the input that might explain why these two participants were primed this way, the finding shows that individuals have their own unique primings of language use. There were more participants in the LIVT group using *is ('s) quite* in comparison with the ones in the other two groups. There were also more participants in the TESOL group using *is ('s) pretty* in comparison with the ones in the other two groups. The finding also shows that more or fewer primed individuals in one group as opposed to another groups is possibly the cause of the significant difference between groups. (See Table 6.35.)

is('s) quite									is('s) pretty								
Participants	Freq.	%	Participants	Freq.	%	Participants	Freq.	%	Participants	Freq.	%	Participants	Freq.	%	Participants	Freq.	%
LIVT	27	100	TE	7	100	NTE	5	100	TE	23	100	LIVT	4	100	NTE	0	0
LIVT03	**12**	**44.44**	TE01	1	14.29	NTE08	5	100	**TE01**	**11**	**47.83**	LIVT04	1	25			
LIVT08	6	22.22	TE08	1	14.29				TE10	6	26.09	LIVT08	3	75			
LIVT11	4	14.81	TE09	1	14.29				TE13	3	13.04						
LIVT01	2	7.41	TE12	4	57.14				TE09	2	8.70						
LIVT04	1	3.70							TE04	1	4.35						
LIVT07	1	3.70															
LIVT12	1	3.70															

Table 6.35 Participants using *is /'s quite* and *is ('s) pretty i*n LIVT, TESOL and NTESOL

6.2.3 Summary of the use of *is*

To test whether the first hypothesis and the second hypothesis are true, we have investigated the use of the auxiliary verb *is* along with its short form *'s* and past tense form *was*. The findings are presented in Table 6.36 and Table 6.37. With regard to our hypothesis 1, we found that the LIVT participants used the collocation *that is ('s)* significantly more frequently than the NTESOL participants. The collocates of *that is ('s)* used by the LIVT participants were similar to the ones used by the speakers in BASEah and COCAsp corpora. Although the BASEah and COCAsp corpora were not the English input that the LIVT participants had received in their studies and daily lives, they represent the types of academic spoken English and semi-formal or casual spoken English that the LIVT participants were very likely to encounter in their one-year living and studying in the UK. The finding supports our first hypothesis that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan. The language input from the immersion of their language envrionments might be responsible for the differences.

Findings that Agree with Hypothesis 1			
Noticeable Differences	Collocations	The types of differences	Descriptions OR noteworthy findings/observation
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	*is quite/ 's quite*	Discourse preference	The LIVT participants used the adverb *quite* to modify adjectives significantly more often than the TESOL and NTESOL participants. There are more participants in the LIVT group using *is ('s) quite* in comparison with those in the other two groups.
The LIVT participants used the collocations **significantly more frequently** than the NTESOL participants.	*that is/ that's*	Lexico-grammatical difference	The collocates of *that is ('s)* used by the LIVT participants are similar to the ones found in the reference corpora, BASEah and COCAsp. The NTESOL participants did not share this similarity.

Table 6.36 Findings in the use of *is* that support Hypothesis 1

We also found that there was another noticeable difference in the LIVT group. The LIVT participants used the collocation *is ('s) quite* significantly more often than the other two groups. Although the significantly frequent use of the collocation *is ('s) quite* in the LIVT group does not provide enough evidence to show that the influence was from the English input they received, the frequent use of one particular expression still shows how people who have had different language backgrounds can behave differently in their language use.

Findings that Agree with Hypothesis 2			
Noticeable Differences	Collocations	The types of differences	The possible factors that affected this difference OR noteworthy observation
The TESOL participants used the collocations **significantly more frequently** than the LIVT and NTESOL participants.	*is pretty/'s pretty*	Discourse preference	The TESOL participants used the adverb *pretty* to modify adjectives significantly more often than the LIVT and NTESOL participants. There were more participants in the TESOL group using *is ('s) pretty* in comparison with those in the other two groups.
The TESOL participants used the collocations **significantly less frequently** than the LIVT and NTESOL participants.	*it's*	Discourse preference	The use of *I think it is/it's because* is a relatively fixed expression when the LIVT and NTESOL participants used the collocation *it is/it's because*. The TESOL participants did not share this tendency.
The TESOL participants used the collocations **significantly more frequently** than the NTESOL participants.	*that is/that's*	Lexico-grammatical Difference	The collocates of *that is ('s)* used by the TESOL participants were similar to the ones found in the reference corpora, BASEah and COCAsp. The NTESOL participants did not share this similarity.
	is that	Lexico-grammatical Difference	The TESOL participants used *is that* in the pattern *NP+ is that + [S]* significantly more often than the LIVT and NTESOL participants.

Table 6.37 Findings in the use of *is* that support Hypothesis 2

As to our second hypothesis, that the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying in non- English-relevant subjects, we found that the TESOL participants used the collocation *that is ('s)* significantly more

often than the NTESOL participants. Also, the way the TESOL participants used this collocation was similar to the way it was used in the reference corpora. The NTESOL participants did not share this similarity. Therefore, this finding supports our second hypothesis. We also found that there was another noticeable difference in the TESOL group. The TESOL participants used the collocation *is ('s) pretty* significantly more frequently than the participants in the other two groups. Although it is difficult to be sure that the significantly frequent use of the collocation *is ('s) pretty* in the TESOL group is the result of the input influence, what we are sure is that the TESOL group had their own uniqueness in language use.

In Chapter 5, we discovered that several combinations of the word 是 [BE] with pronouns (是 我 [BE I], 是 你 [BE you], 我 是 [I BE] and 它 是[it BE]) were used by one group of participants significantly more or less often than by the other groups of participants. These findings challenge our null hypotheses 3 and 4. Therefore, in the next section, we will investigate the use of Mandarin word 是 (shì) in detail.

6.3 The use of *是(shì) [BE]*

In this section, we will investigate the use of Mandarin word 是 (shì) in order to test our third and fourth hypotheses:

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

Beginning by looking at the frequency of use of 是 (shì) across the three groups, it is found that the word 是 (shì) was used significantly more frequently by the LIVT participants in comparison with the TESOL participants, although the significance is weak. The use of 是 (shì) in the LIVT data is 29.22 times per 1000 words, and the use of 是 (shì) in the TESOL data is 27.04 times per 1000 words. (See Table 6.38.)

		Freq.	Per 1000 words		LLR	Sig.		
是(shì)	LIVT	1127	29.22	**LIVT-TE**	**4.368**	**0.037**	*****	**+**
	TESOL	941	26.65	LIVT-NTE	3.399	0.065		+
	NTESOL	1139	27.04	TE-NTE	0.111	0.739		-

Table 6.38 Log-Likelihood Ratios: 是 (shì) in LIVT, TESOL and NTESOL

It should be noted that in Mandarin Chinese, the word 是 (shì) is often used as a morpheme (suffix) in adverbs or connectives or even prepositions, such as the connectives 可是 (which means 'but') and 或是 [or], the adverbs 就是 [just], 还是 (usually 'still') and 只是 [only/simply], and the preposition 像是 [like/as]. (See Table 6.39.) The tool used for tokenization in this study treated

the use of 是 (shì) in these terms and the word 是 (shì) on its own as different entries[10]. We found that some of these terms were used significantly more frequently by one group of participants than the other groups of participants. For instance, the LIVT participants used the term 还是 (which means 'still') significantly more often than the TESOL and NTESOL participants. The LIVT participants also used the term 像是 (which means 'like/as') significantly less often than the TESOL and NTESOL participants. (See Table 6.40.) Although it would be worth studying further these differences in the use of connectives, adverbs or prepositions because of these significant differences, to do so would take us beyond our concern with the main use of 是(shì) in this study. In this chapter, we simply focus on how the word 是(shì) is used as a single unit.

	LIVT			TESOL			NTESOL		
N	Words	Freq.	per 1000 words	Words	Freq.	per 1000 words	Words	Freq.	per 1000 words
1	就是	1040	26.96	就是	910	25.77	就是	1,202	28.54
2	可是	190	4.93	可是	189	5.35	可是	167	3.96
3	但是	134	3.47	但是	114	3.23	但是	165	3.92
4	还是	131	3.40	还是	81	2.29	还是	104	2.47
5	或是	69	1.79	或是	80	2.27	或是	97	2.30
6	只是	56	1.45	只是	36	1.02	只是	50	1.19
7	要是	11	0.29	像是	28	0.79	像是	39	0.93
8	算是	11	0.29	就是说	20	0.57	是也	17	0.40
9	就是说	9	0.23	算是	14	0.40	算是	12	0.28
10	像是	9	0.23	是也	12	0.34	就是说	8	0.19
11				要是	9	0.25	于是	6	0.14
12							总是	5	0.12

Table 6.39 Use of *是 (shì)in LIVT, TESOL and NTESOL

[10] Although the tool tried to separate the connectives and adverbs containing 是 (shì) from the main use of 是 (shì), sometimes the separation was quite difficult to make. For instance, there is a term 或者 in Mandarin that can combine with 是 (shì) as one unit. The phrase 或者是 functions as a connective similar to 或是 meaning 'or'. However, the term 或者 is also commonly used alone (also meaning 'or') or used in other combinations, such as 或者说 or 或者是说 (all meaning 'or'). The difference between these phrases is the degree of 'action' and 'emphasis' that the speaker adds to his/her speech. For instance, the term 或者说 is similar to the meaning 'or (someone) say that…'. The term 或者是说 is similar to the meaning 'or this is to say that...' Mandarin Chinese utilises this method of composition greatly to convey different meanings subtly. Therefore, the clear separation and categorisation of these uses of 是 (shì) is really difficult. One may argue that the word 或 in the term 或是 can also be used alone (both mean 'or'). However, since the SINICA tokenization tool sees the term 或者 as the L1 collocate of 是 and treats the use of 或是 as one entry, for convenience we will apply the same principle of division.

LIVT - NTE					LIVT - TE					TE - NTE				
Over					**Over**					**Over**				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
要是	*7.677*	*0.006*	**	+	**还是**	**7.907**	**0.005**	**	+	*可是*	*8.000*	*0.005*	**	+
还是	**5.942**	**0.015**	*	+						就是说	7.646	0.006	**	+
可是	*4.201*	*0.040*	*	+	**Less**					*要是*	*6.137*	*0.013*	*	+
Less					Words	LLR	Sig.			**Less**				
Words	LLR	Sig.			像是	11.986	0.001	***	-	Words	LLR	Sig.		
像是	17.666	0.000	***	-	就是说	5.305	0.021	*	-	就是	5.415	0.020	*	-

Table 6.40 Log-Likelihood Ratios: *是 (shì) in LIVT, TESOL and NTESOL

6.3.1 L1 Collocates of 是 (shì) [BE]

Table 6.41 presents the top 10 most frequent L1 collocates of the word 是 (shì). We can see that there were five collocates shared by all the three groups. The negative marker 不, the adverbs 也 [also] and 都 [all/both/entirely], the term 或者 [‘or’ or perhaps] and the term 应该 [ought to/should] were commonly used by all the three groups. There were however several collocates occurring only in two groups’ top 10 most frequent collocate lists. For instance, the term 如果 (means ‘if’) appears in the TESOL and NTESOL groups’ lists, but not in LIVT’s list. The verb 觉得 [think] appears in the LIVT and NTESOL groups’ lists, but not in TESOL’s list. We will examine these use via log-likelihood ratios.

	LIVT	1127	29.22	TESOL	941	26.65	NTESOL	1139	27.04
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	不	116	3.01	不	93	2.63	不	121	2.87
2	也	82	2.13	或者	64	1.81	或者	80	1.90
3	都	65	1.69	都	55	1.56	也	67	1.59
4	或者	62	1.61	也	49	1.39	都	62	1.47
5	觉得	44	1.14	如果	43	1.22	它	51	1.21
6	我	42	1.09	我	37	1.05	觉得	33	0.78
7	可能	29	0.75	的	27	0.76	如果	32	0.76
8	应该	28	0.73	应该	27	0.76	那	27	0.64
9	的	25	0.65	这	25	0.71	应该	26	0.62
10	那	21	0.54	它	23	0.65	可能	26	0.62

Table 6.41 L1 Collocates of 是 (shì) in LIVT, TESOL and NTESOL

LIVT - NTE						**LIVT - TE**						**TE - NTE**					
Over						**Over**						**Over**					
Word	With	LLR	Sig.			Word	With	LLR	Sig.			Word	With	LLR	Sig.		
我	**是**	**10.037**	**0.002**	****	+	觉得	是	18.398	0.000	***	+	**我**	**是**	**8.587**	**0.003**	****	+
Less						也	是	5.270	0.022	*	+	如果	是	4.140	0.042	*	+
Word	With	LLR	Sig.			**Less**						**Less**					
它	**是**	**11.842**	**0.001**	****	-	Word	With	LLR	Sig.			Word	With	LLR	Sig.		
如果	是	5.739	0.017	*	-	如果	是	18.136	0.000	***	-	觉得	是	7.340	0.007	**	-
												它	**是**	**6.035**	**0.014**	*	-
												那	是	5.402	0.020	*	-

Table 6.42 Log-Likelihood Ratios: 是 (shì) in LIVT, TESOL and NTESOL – L1 Collocates

In Table 6.42, we can see that the collocation 我是 [I BE] was used significantly more often by the LIVT and TESOL participants than by the NTESOL participants. The NTESOL participants, however, used the collocation 它是 [it BE] significantly more frequently than the LIVT and TESOL participants. The LIVT and NTESOL participants used the collocation 觉得是 [think BE] significantly more frequently than the TESOL participants. The LIVT participants used the collocation 如果是 [if BE] significantly less often than the TESOL and NTESOL participants. All of

these findings indicate that our null hypotheses are incorrect. From the observation on the instances of 觉得是 [think BE] in the groups, we found that this collocation often occurrs in the bigger collocation 我 觉得 是 [I think BE]. This is similar to the English expression *I think* and was discussed in chapter 5.2.1 and 5.2.2. Therefore, we will not further investigate the use of 觉得是 [think BE]. The collocation 如果是 [if BE] is used as a connective in Mandarin. As we have mentioned, although it would have great value to see how these EFL learners use connectives differently, we have not chosen to include them in our investigation in the current study. Instead, we will investigate the use of 我是 [I BE] and 它是 [it BE] for testing our hypotheses.

我是 (shì) [I BE]

The collocation 我是[I BE] was used significantly more frequently by the LIVT and TESOL participants in comparison with the NTESOL participants. (See Table 6.43.) There are 42 occurrences of the collocation 我是 [I BE] in the LIVT data (1.09 times per 1000 words), and 36 occurrences of 我是 in the TESOL data (1.02 times per 1000 words). There are 19 occurrences of the collocation 我是 in the NTESOL data (0.45 times per 1000 words.) The LIVT and TESOL participants used the collocation 我是 [I BE] about twice as often as the NTESOL participants. (See Table 6.44)

我 是(shì)	**LLR**	**Sig.**		
LIVT - TE	0.029	0.864		+
LIVT - NTE	**10.037**	**0.002**	****	**+**
TE - NTE	**8.587**	**0.003**	****	**+**
是(shì)	**LLR**	**Sig.**		
LIVT - TE	4.368	0.037	*	+
LIVT - NTE	3.399	0.065		+
TE – NTE	0.111	0.739		-

Table 6.43 Log-Likelihood Ratios: 我是 (shì)and 是(shì)

	LIVT			TESOL			NTESOL		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
是 (shì)	1127	100	29.22	941	100	26.65	1139	100	27.04
我是 (wǒ shì)	42	3.73	**1.09**	36	3.83	**1.02**	19	1.67	**0.45**
Patterns of 我 是 (wǒ shì)	N L1 Centre R1 R2 1 可是 我 是 觉 得 觉 得 kěshì wǒ shì juédé juédé but I + SHI think think 2 因 为 这 麽 知 道 yīnwèi zhème zhīdào because so/this wa know 3 觉 得 因 为 juédé yīnwèi think because			N L1 Centre R1 R2 1 觉 得 我 是 觉 得 觉 得 juédé wǒ shì juédé juédé think I + SHI think think 2 因 为 yīnwèi because 3 但 是 dànshì but 4 然 後 ránhòu then			N L1 Centre R1 1 我 是 觉 得 wǒ shì juédé I + SHI think 2 这 样 zhèyàng so/this way		

Table 6.44 Use of *是 (shì)* and 我是(wǒ) (shì) in LIVT, TESOL and NTESOL

When observing the patterns of the collocation 我是 (wǒ shì) [I BE], it is found that all the participants tended to use 觉得 (juédé) [think] after 我是 (wǒ shì) [I BE]. There are 7 occurrences of 我是觉得 (wǒ shì juédé) in the LIVT data (0.18 times per 1000 words), 9 occurrences of 我是觉得 (wǒ shì juédé) in the TESOL data (0.25 times per 1000 words) and 6 occurrences of 我是觉得 (wǒ shì juédé) in the NTESOL data (0.14 times per 1000 words). The use of 我是觉得 (wǒ shì juédé) [I BE think] is a variety of 我 觉得 (wǒ juédé) [I think]. This use has a function similar to the use of *I think* in English, and the word 是 (shì) [BE] is used for adding an extra discourse use of emphasis to the action.

Before we explore these differences further, there is a need to say something more about the use of 是 (shì) in Mandarin.

In Ross and Ma's book 'Modern Mandarin Chinese Grammar' (2006, 2014: 61-63), they introduce the word 是 (shì) as a equational verb, and list several features of its usage. The word 是 (shì) connects two noun phrases and implies that these two noun phrases have an equational relationship. For instance,

他 是 学 生 。
Tā shì xuésheng。
He is a student.

The word 是 (shì) is not often used with adjectival verbs or stative verbs. If the word is used with adjectival verbs or stative verbs, it is for special emphasis. This kind of emphasis often has contrastive meaning. Example (A) shows that in Mandarin a speaker does not really need to use the word 是(shì) in a sentence where a sentence of equivalent meaning in English requires a BE-verb. Example (B) illustrates a situation where the addition of the word 是(shì) has a particular discourse function.

(A)	他 很 聪 明 。
	Tā hěn cōngmíng。
	He [is] very intelligent.
(B)	他 **是** 很 聪 明 。
	Tā **shì** hěn cōngmíng。
	He IS very intelligent (despite what you may think.)

When the complement of 是 (shì) includes a number, the word 是(shì) can be omitted in an affirmative sentence. (It cannot be omitted when the object is negated.)

(C)	我 弟弟 **(是)** 十 岁 。
	Wǒ dìdi (shì) shísuì。
	My brother [is] ten years old.
(D)	我 弟弟 **不是** 十 岁 。
	Wǒ dìdi búshì shísuì。
	My brother **is not** ten years old.

The word 是 (shì) is also used for adding focus on some details of a situation. The focus can be on the time (Example E), place (Example F), or the material that a thing is made of (Example G).

(E)	他 **是** 今 天 早 上 来 的。
	Tā **shì** jīntiān zǎoshɑng lai de。
	He came this morning. (It was this morning that he came.)
(F)	我 **是** 在 学 校 认 识 他 的。
	Wǒ **shì** zài xuéxiào rènshí tā de。
	I met him at school. (It was at school where I met him.)
(G)	这 个 杯 子 **是** 木头 做 的。
	Zhè ge bēizi **shì** mùtou zuò de。
	This cup is made of wood.

As a native Mandarin speaker myself, I would like to add to this description. The word 是(shì) in Example (E) and Example (F) can be omitted and the sentences still sound natural to a native speaker, but they lose the sense of emphasis. (See Example (E) and Example (F). Example (G), however, will sound a little strange to a native speaker's ears if the word 是 (shì) is omitted. (See Example (G).) It can still be understood, but there is a little oddness if 是 (shì) is omitted where it would normally be used to focus on the material that one thing is made of. To take another example, there is a saying 女人是水做的 (Women are made of water.) If we took out the 是 (shì), it would seem a very strange sentence to a Mandarin speaker. Therefore, in this situation, the word 是 (shì) is seen as a necessary use.

(E)	他 今 天 早 上 来 的。
	Tā jīntiān zǎoshɑng lai de。
	He came this morning.
(F]	我 在 学 校 认 识 他 的。
	Wǒ zài xuéxiào rènshí tā de。
	I met him at school.
(G]	*这 个 杯 子 木头 做 的。
	Zhè ge bēizi mùtou zuò de。
	This cup is made of wood.
(H)	女人 **是** 水 做 的。 /* 女人 水 做 的。
	Nǚrén **shì** shuǐ zuò de。 /* Nǚrén shuǐ zuò de。
	Women are made of water.

Previously, we checked the use of 我是 (wǒshì) [I BE] in all the three groups and found that the most frequent collocation of 我是 (wǒshì) [I BE] was 我是觉得 (wǒshìjuéde) [I BE think]. The use of 是 (shì) here with 觉得 (juéde) [think] adds emphasis to the common use of 我觉得 (wǒjuéde) [I think] as shown in the examples below.

(I)	我 觉 得 他 很 和善 。
	Wǒ juédé tā hěn héshàn
	I think he is very kind.
(J)	我 **是** 觉 得 他 很 和善 。
	Wǒ **shì** juédé tā hěn héshàn。
	I think he is very kind (despite what you think of him.)

The three groups of participants showed no significant differences in terms of the occurrences of the emphasis 是 (shì) with the verb 觉得 [think]. There are 0.29 instances per 1000 words of 是觉得 [BE think] and 0.18 instances per 1000 words of 我是觉得 [I BE think] in the LIVT data. There are 0.34 instances per 1000 words of 是觉得 [BE think] and 0.25 instances per 1000 words of 我是觉得 [I BE think] in the TESOL data. There are 0.24 instances per 1000 words of 是觉得 [BE think] and 0.14 instances of 我是觉得 [I BE think] in the NTESOL data. (See Table 6.45.) The frequencies suggest the Taiwanese participants did not differ greatly in how they added emphasis 是 (shì) to the verb 觉得 [think]. As far as the use of the collocate 觉得 [think] is concerned, the findings show that there were no noticeable differences between the three groups with regard to their lexico-grammatical usage. However, as already noted, we did find a noticeable difference in the LIVT and TESOL groups. The LIVT and TESOL participants used the collocation 我是 [I BE] more than twice as often as the NTESOL participants. Since the most frequent collocate 觉得 [think] cannot provide us the possible explanation to the difference, we move on to view the total instances of 我是 [I BE] in the three groups.

	LIVT		**TESOL**		**NTESOL**	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
觉得 (juéde)	685	17.76	559	15.83	781	18.54
是 (shì) 觉得 (juéde)	11	**0.29**	12	**0.34**	10	**0.24**
我(wǒ)觉得 (juéde)	487	12.63	377	10.68	521	12.37
我(wǒ) 是 (shì) 觉得 (juéde)	7	**0.18**	9	**0.25**	6	**0.14**

Table 6.45 Use of *觉得 (juéde), 是 (shì) 觉得 (juéde), 我(wǒ)觉得 (juéde)* and 我(wǒ) 是 (shì) 觉得 (juéde) in LIVT, TESOL and NTESOL

When investigating all the instances of the collocation 我是 in the three groups, it is found that the actual situations in which the use of the word 是 (shì) is necessary are rare. There are four instances of 我是 [I BE] that the use of 是 (shì) is grammatically required in the LIVT group. There are 9 instances of this kind of use in the TESOL group. There is only one instance of this kind of use

in the NTESOL group. A large part of the remaining instances used the word 是 (shì) for its emphasis function. There were 33 instances in the LIVT group using 是 (shì) for emphasis (0.86 times per 1000 words.) There were 36 instances in the TESOL group using 是(shì) for emphasis (0.71 times per 1000 words.) There were only 17 instances in the NTESOL group using 是 (shì) for emphasis (0.40 times per 1000 words.) The LIVT and TESOL participants used the word 是 (shì) for emphasis purpose twice as often as the NTESOL participants. (See Table 6.46.) This finding provides us with a possible explanation of the significantly high frequency of the use of 我是 (wǒ shì) in the LIVT and TESOL groups in comparison to the NTESOL group. That is, their discourse preferences were different. The LIVT and TESOL participants tended to add emphasis on the following verbs with the word 是 (shì) when the subjects were themselves, while the NTESOL participants did not.

	LIVT		**TESOL**		**NTESOL**	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
是(shì)	1127	29.22	941	26.65	1139	27.04
我是 (wǒ shì)	42	1.09	36	1.02	19	0.45
是 (shì) used for emphasis	33	**0.86**	25	**0.71**	17	**0.40**
是 (shì) used in necessity	4	**0.10**	9	**0.25**	1	**0.02**
Cannot be classified[11]	5	0.13	2	0.06	1	0.02

Table 6.46 Use of 我是 (wǒ shì) in LIVT, TESOL and NTESOL

What we have found suggests that the differences of the use of 我是 (wǒ shì) [I BE] between the groups are likely to be a result of their different discourse preferences. However, it is difficult to know why the LIVT and TESOL participants should have be primed for such a preference in their use of 是 (shì) in a particular context more strongly than the NTESOL participants. The statistical significance suggests that the differences of frequency are not a coincidence, so it indicates that there must be some factors affecting these participants' Mandarin use. However, the LIVT, TESOL and NTESOL participants should have had no significantly distinct differences in their Mandarin input, since they

[11] These instances are not fully constructed sentences. The participants dropped words and started another sentence.

were all from the same Mandarin-speaking environment. They were all postgraduate students, so they were of similar social status/level of education. They were asked identical questions under the same circumstances and were interviewed individually by the same interviewer. The only difference between these groups is the different level of English input (studying abroad/English-relevant subject) to which they had been exposed. As with what we found in chapter 5, it is possible that there exists a cross-language influence on the LIVT and TESOL participants' English and Mandarin use. The English BE-verbs do not have the same discourse use as the Mandarin word 是 [BE] does. However, in English, words for expressing modality/modification (a discourse strategy) often occur directly after pronouns. The frequent use of words for emphasis (another discourse strategy) (e.g. 是 [BE]) following directly after the pronoun 我 [I] in the LIVT and TESOL groups is, perhaps, influenced by the LIVT and TESOL participants' relatively more frequent use of English (which has this kind of feature.) Of course, it will need more research and evidence to test this possible explanation. Nevertheless, what we found in the use of 我 是 [I *BE*] does show how different the LIVT and TESOL participants were from the NTESOL participants in their Mandarin use. Hence, hypotheses 3 and 4 are proved to be incorrect. Next, we move to investigate the use of 它是 [it BE].

它是 (tā shì) [it BE]

With regard to the use of the collocation 它是 [it BE], it is found that the LIVT and TESOL participants used 它是 [it BE] significantly less often than the NTESOL participants. (See Table 6.47.) There are 51 instances of 它是 [it BE] in the NTESOL data (1.21 times per 1000 words.) In the LIVT data, the collocation 它是 [it BE] occurs 19 times (0.49 times per 1000 words.) It occurs 23 times in the TESOL data (0.65 times per 1000 words.) The LIVT and TESOL participants used the collocation 它是 [it BE] about half as frequently as the NTESOL group. (See Table 6.48.) The significantly low use of 它是 [it BE] in the LIVT and TESOL groups is very likely the result of the infrequent use of the impersonal pronoun 它 [it] as we have discussed in chapter 4. Next, we will investigate how these groups used 它是 [it BE] in detail.

它是 (tā shì)	**LLR**	**Sig.**		
LIVT - TE	0.816	0.366		-
LIVT - NTE	11.842	**0.001**	***	-
TE - NTE	6.035	**0.014**	*	-
它 (tā)	**LLR**	**Sig.**		
LIVT-TE	0.067	0.796		+
LIVT-NTE	26.400	**0.000**	***	-
TE-NTE	27.762	**0.000**	***	-

Table 6.47 Log-Likelihood Ratios: 它是 (tā shì)

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是 (shì)	1127	100	29.22	941	100	26.65	1139	100	27.04
它是 (tā shì)	19	1.69	**0.49**	23	2.44	**0.65**	51	4.48	**1.21**
Patterns of *它是 (tā shì)*									

N	L1	Centre	R1
1	就 是	它 是	因 为
	jiùshì	tā shì	yīnwèi
	just	it + SHI	because
2	因 为		
	yīnwèi		
	because		

N	L1	Centre	R1
1	因 为	它 是	
	yīnwèi	tā shì	
	because	it + SHI	

N	L1	Centre	R1
1	觉 得	它 是	比较
	juédé	tā shì	bǐjiào
	think	it + SHI	compare
2	因 为		就 是
	yīnwèi		jiùshì
	because		just
3	简体字		
	jiǎntǐzì		
	simplified word		
4	就 是		
	jiùshì		
	just		

Table 6.48 Use of *是(shì)* and *它是(tāshì)* in LIVT, TESOL and NTESOL (1)

In Table 6.48, we do not see too much difference in terms of the patterns of the use of 它是[tāshì] [it BE]. The collocation 它是[tāshì] [it BE] is usually surrounded by the connective 因为[yīnwéi] [because] or adverb 就是[jiùshì] [just]. However, when examining the instances of the collocation 它是[tāshì][it BE] with the same method that we used for 我是[wǒshì][I BE], the reason that caused the distinct differences between

groups is found. Before we go into the potential explanation to the subject, there is a need to illustrate how the collocation 它是[it BE] is usually used in spoken Mandarin.

The word 是 (shì), as mentioned previously, is usually used to connect two noun phrases. It does not have to be used with adjectival verbs (e.g. 漂亮 [beautiful]) or stative verbs (e.g. 累 [tired] or 饿 [hungry]), but it can be used for expressing emphasis. The word 是(shì) is used to add focus on some details of a situation or to lead into the feature/quality/nature of a subject. In the former use, the word 是(shì) can be omitted without violating a native speaker's intuition, but it will lose a sense of emphasis in the sentence. In the later use, it would sound a little odd to a native speaker if the word 是 (shì) is omitted. One may argue that such an expression is still understandable and can be used by some speakers if they want, but in this study we still see the use of 是 (shì) in this kind of expression as a necessary element. For instance, Example NTE6-1 demonstrates when a speaker illustrates a feature of a thing. The use of the word 是(shì) is required. (The subject in Example NTE6-1 is modern poetry.) The NTESOL participant was saying that in comparison to the classic poems it is much easier for someone to create a modern poem. S/he attributed to modern poems the feature of being 'easy to be created.' In this situation, the use of 是 (shì) is necessary.

NTE6-1	因 为 它 [新 诗] **是** 比较 容 易 去 , 去 创 作 的。
	Yīnwèi tā [xīnshī] **shì** bǐjiào róngyì qù , qù chuàngzuò de。
	Because it [a modern poem.] is easier (for someone) to create/write.
	*****因 为 它 [新 诗] 比较 容 易 去 , 去 创 作 的。
	Yīnwèi tā [xīnshī] bǐjiào róngyì qù , qù chuàngzuò de。

Example TE6-1 below demonstrates how the word 是 (shì) connects two noun phrases. The pronoun 它 refers to 'writing', and the TESOL participant was saying that writing equals the process of thinking. In this situation, the use of 是 (shì) is necessary.

TE6-1	它 [写 作] **是** 一 种 帮 助 思考 的 过 程 。
	Tā [xiězuò] **shì** yī zhǒng bāngzhù sīkǎo de guòchéng。
	It [writing] is a process that assists (someone) thinking.
	*****它 [写 作] 一 种 帮 助 思考 的 过 程 。
	Tā [xiězuò] yī zhǒng bāngzhù sīkǎo de guòchéng。

Example LIVT6-1 demonstrates how 是 (shì) is used for emphasis. The LIVT participant was talking about Chinese characters. There are two kinds of Chinese characters in writing: traditional Chinese characters (e.g. 龍 lóng [dragon]), and simplified Chinese characters (e.g. 龙 lóng

[dragon]). In Taiwan, we still use traditional Chinese characters in writing. In the Example LIVT6-1, the LIVT participant added a little emphasis on the place where traditional characters are used. In this situation, the use of the word 是 (shì) is not necessary and can be omitted.

LIVT6-1	它 [繁 体字] **是** 在 台 湾 已经 被 使 用 很 久 了。
	Tā [fántǐzì] **shì** zài táiwān yǐjīng bèi shǐyòng hěn jiǔ le。
	It [Traditional Chinese characters] has been used in Taiwan for a long time.
	它 [繁 体字] 在 台 湾 已经 被 使 用 很 久 了。
	Tā [fántǐzì] zài táiwān yǐjīng bèi shǐyòng hěn jiǔ le。

When investigating all the instances of the use of 它是 [it BE] in the three groups, it is found that the NTESOL participants used 它是 in the situations where the use of 是 (shì) is a necessity more often than the LIVT and TESOL participants. This kind of use occurs 31 times (0.74 times per 1000 words) in the NTESOL data. It has only 11 instances of this use in the LIVT data (0.29 times per 1000 words), and 16 instances of this use in the TESOL data (0.45 times per 1000 words.) The NTESOL participants had this kind of use in their spoken Mandarin about twice as often as the LIVT and TESOL participants. The NTESOL participants also used 是 (shì) in the collocation 它是 [it BE] for the purpose of emphasis more often than the LIVT and TESOL participants. There were 14 occurrences of emphasis 是 (shì) in the use of 它是 [it BE] in the NTESOL data (0.33 times per 1000 words.) There are only 5 instances of emphasis 是 (shì) in 它是 [it BE] (0.13 times per 1000 words) in the LIVT data and 6 instances of emphasis 是(shì) in 它是 [it BE] (0.17 times per 1000 words) in the TESOL data. The NTESOL participants used emphasis 是 (shì) in 它是 [it BE] about twice as often as the LIVT and TESOL participants. (See Table 6.49.)

These findings show why there is a significant difference between the LIVT and TESOL groups and the NTESOL group in their use of 它是 [it BE]. Previously we found that the LIVT and TESOL participants tended to use emphasis 是 (shì) to add extra discourse meaning to what they thought or did. They also used 是(shì) for connecting the subject (themselves) with a noun phrase to illustrate their conditions. What we have found with regard to the use of 它是 [it be] is an opposite situation. The LIVT and TESOL participants What we have found with regard to the use of 它是 [it be] is an opposite situation. The LIVT and TESOL participants did not use it as often as the NTESOL participants. The LIVT and TESOL participants tended not to use the pronoun 它 [it] when describing something's features or qualities or characterising it with a noun phrase as often as the

NTESOL participants. It suggests that the LIVT and TESOL participants may have been primed differently for the use of pronoun 它 [it] in comparison with the NTESOL participants. The NTESOL participants tended to use the impersonal pronoun 它 [it] to refer to a precise referent. The LIVT and TESOSL participants, on the other hand, were not primed to use 它 [it] to refer to a precise referent as strongly as the NTESOL participants. This difference, perhaps, can again be explained by their different English input, as discussed in connection with the use of 我是[I BE]. In English, the collocation *it is (it's)* is used in a complex range of structures. The use of *it* as a pronoun with a precise referent is not frequently seen. (In this situation, the demonstrators *this* and *that* are more likely to be used.) Therefore, the LIVT and TESOL participants, who had been exposed to more English input and have had the opportunities to practice English more than the NTESOL participants, were potentially influenced by the English priming for *it* in their Mandarin use of 它 [it].

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是 (shì)	1127	--	29.22	941	--	26.65	1139	--	27.04
它是 (tā shì)	19	100	0.49	23	100	0.65	51	100	1.21
是 (shì) used for emphasis	5	26.32	**0.13**	6	26.09	**0.17**	14	27.45	**0.33**
是 (shì) used in necessity	11	57.89	**0.29**	16	69.57	**0.45**	31	60.78	**0.74**
Cannot be classified	3	15.79	0.08	1	4.35	0.03	6	11.76	0.14
是 (shì) used in necessity	11	100	0.29	16	100	0.45	31	100	0.74
是 (shì) used in necessity with 的(de) [12]	1	9.09	0.03	1	6.25	0.03	8	25.81	0.19

Table 6.49 Use of *是 (shì)* and *它是 (tā shì)* in LIVT, TESOL and NTESOL (2)

The significant differences that we found in the use of 我是 [I BE] and 它是 [it BE] amongst these groups challenge our third and fourth hypotheses. These participants were born in Taiwan, had lived in Taiwan most of their lives and had received similar Mandarin language input in

[12] When the word 是 is used in the pattern 是+[adjectival verb] or [有+ noun]+的, the word 是 cannot be omitted. This expression is used to describe the feature of the subject. For instance, 它是自然的。[*它自然的。] = It is natural. 它是畅销的。[*它畅销的。] = It is best-selling. Example NTE6-1 is a complex form of such use.

their daily lives and their previous formal school education. In our interviews, as noted earlier, they were asked the same set of questions. Of course the content of their answers to the questions were different, but there is no reason to suppose that differences in content should arise between groups such that the differences in their answers would influence how they used the word 是 (shì) with the pronouns 我 [I] and 它[it]. Our findings suggest that perhaps there are other factors influencing how these participants used their spoken Mandarin, and we propose that the differences in the extent and nature of their English input and English use is one of the possible factors. Next, we will investigate the R1 collocates of 是(shì).

6.3.2 R1 Collocates of 是(shì) [BE]

Table 6.50 presents the top 10 most frequent R1 collocates of 是 (shì) in the three groups. The shared R1 collocates are the verb 有 [to have], the adverb 很 [very], the number 一 [one], the pronoun 我 [I], the connective 因为 [because] and the negative marker 不 [no/not]. Some words are found in two of the groups' top 10 most frequent R1 collocate lists. The verb 说 [say] and the pronoun 你 [you] are found in the LIVT and TESOL groups' lists. The word 要 [want/must/going to] is found in the TESOL and NTESOL groups' lists. The determiner 那 [that] and the adverb 就是 [just] are found in the LIVT and NTESOL groups' lists. The adverb 什么 [what] only appears in the NTESOL group's list.

	LIVT	1127	29.22	TESOL	941	26.65	NTESOL	1139	27.04
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	有	61	1.58	有	48	1.36	一	71	1.69
2	很	53	1.37	很	38	1.08	很	59	1.40
3	一	50	1.30	一	38	1.08	有	52	1.23
4	我	46	1.19	什么	30	0.85	因为	40	0.95
5	你	45	1.17	我	29	0.82	要	32	0.76
6	说	35	0.91	你	29	0.82	就是	31	0.74
7	不	34	0.88	说	21	0.59	我	30	0.71
8	那	32	0.83	因为	19	0.54	不	30	0.71
9	因为	31	0.80	不	18	0.51	什么	29	0.69
10	就是	26	0.67	要	17	0.48	那	20	0.47

Table 6.50 R1 Collocates of 是 (shì) in LIVT, TESOL and NTESOL

When checking these collocates with log-likelihood ratios (See Table 6.51), it is found that the collocation 是 你 [BE you] was used significantly more often by the LIVT and TESOL participants than by the NTESOL participants. The collocations 是 我[BE I] and 是 那[BE that] were used significantly more often by the LIVT participants than by the TESOL and NTESOL participants.

There are also some other collocations, such as 是 就 是 [BE just], 是一 [BE one], 是 因为 [BE because], 是 不 [BE not/no] and 是 什 么 [BE what], which show weak significance. We will not however include them in our further investigation because we would like to keep our focus on the use of the combinations of pronouns and the word 是 (是 你 [BE you] and 是 我 [BE I]). In our earlier investigation on the use of 你 [you] (in 5.3.2), we found that the LIVT and TESOL participants used the verb 说 [say] as a R1 collocate of 你 (你说 [you say]) significantly more often than the NTESOL participants. Here we find again that the LIVT participants used the verb 说 with 是 (是 说[BE say]) significantly more often than the NTESOL participants. The LIVT participants used certain collocations containing the word 说 [say] more often than the participants of the other groups, and this difference is worth studying. Therefore, we will include the collocation 是说 [BE say] in our further investigation. In the following section, we will investigate the use of the collocations 是 你 [BE you], 是 我 [BE I] and 是 说 [BE say].

LIVT - NTE						**LIVT - TE**						**TE - NTE**					
Over						Over						Over					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
是	**你**	21.613	**0.000**	***	+	**是**	**我**	**9.511**	**0.002**	**	+	**是**	**你**	8.579	**0.003**	**	+
是	**我**	14.363	**0.000**	***	+	是	那	6.691	0.010	**	+	Less					
是	**说**	6.945	**0.008**	**	+							With	Word	LLR	Sig.		
是	什么	5.530	0.019	*	+							是	一	5.005	0.025	*	-
是	那	4.594	0.032	*	+							是	就是	4.753	0.029	*	-
												是	因为	4.397	0.036	*	-
												是	不	4.245	0.039	*	-

Table 6.51 Log-Likelihood Ratios: *是(shì)* in LIVT, TESOL and NTESOL– R1 Collocates

是你 (shì nǐ)[BE you]

The LIVT and TESOL participants used the collocation 是 你 [BE you] more often than the NTESOL participants. (See Table 6.52.) There are 25 instances of 是 你[BE you] (0.71 times per 1000 words) found in the TESOL group, 17 instances of 是 你[BE you] (0.44 times per 1000 words) found in the LIVT group and only 9 instances of 是 你[BE you] (0.21 times per 1000 words) found in the NTESOL group Although the TESOL and LIVT participants respectively used the collocation 是 你[BE you] twice and three times more often than the NTESOL participants, from the observation on the instances, there is no noticeable difference in terms of the way the collocation was used found. (See Table 6.53.)

是你 (shì nǐ)	LLR	Sig.		
LIVT - TE	2.638	0.104		+
LIVT - NTE	21.613	**0.000**	***	+
TE - NTE	8.579	**0.003**	**	+

Table 6.52 Log-Likelihood Ratios: *是你(shìnǐ)*

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是 (shì)	1127	100	29.22	941	100	26.65	1139	100	27.04
是你 (shì nǐ)	17	1.51	**0.44**	25	2.66	**0.71**	9	0.79	**0.21**
Patterns of *是你 (shì nǐ)*	N / L1 / Centre / R1 1 / 或 者 / 是 你 / 可能 / huòzhě / shì nǐ / kěnéng / or/perhaps / SHI+YOU / maybe			N / L1 / Centre / R1 1 / 或 者 / 是 你 / 可以 / huòzhě / shì nǐ / kéyǐ / or/perhaps / SHI+YOU / can 2 / 问 题 / wèntí / problem			N / L1 / Centre / R1 1 / 或 者 / 是 你 / huòzhě / shì nǐ / or/perhaps / SHI+YOU		

Table 6.53 Use of 是(shì) and 是你(shìnǐ) in LIVT, TESOL and NTESOL

The most frequent collocate of 是 你[BE you] in the three groups is the adverb/connective 或者 [perhaps/or]. In Table 6.53, it is seen that the patterns of 是 你[BE you] is the same[13] in all the three groups. Previously we mentioned in the footnote 6 that the word 是 (shì) is frequently used as a suffix to the term 或者[perhaps/or] to be an adverb or a connective. The suffix 是 (shì) used in the adverbs or connectives attaches the adverbial words/connective words to the following clauses. Therefore, it is often used at the beginning of a sentence. (See Example TE6-2) There are 14 occurrences of suffix 是 (shì) in the collocation 是 你[BE you] found in the TESOL data (0.40 times per 1000 words), 13 occurrences of suffix 是 (shì) in the collocation 是 你[BE you] found in the LIVT data (0.34 time per 1000 words) and only 4 occurrences of suffix 是 (shì) in the collocation 是 你[BE you] found in the NTESOL data (0.09 times per 1000 words.) (See Table 6.54.) The LIVT and TESOL participants used the suffix 是 (shì) in the use of 是 你[BE you] with similar frequency, but this combination was not used very frequently by the NTESOL participants.

TE6-2	或 者 **是** 你 真 的 要 了 解 你 才 写 得 出 东 西 来
	Huòzhě **shì** nǐ zhēnde yào liáojiě nǐ cái xiě dé chū dōngxi lái
	Or perhaps you really have to understand [your topic] so that you can write something [about it.]

The word 是 (shì), as we discussed in the previous section, can also serve as an emphasis unit. This kind of use is however rarely found in our instances of 是 你 [BE you]. There is only one instance of emphasis 是 (shì) found in the LIVT data, one instance of emphasis 是 (shì) found in the NTESOL data and none at all in the TESOL data. The word 是 (shì) also functions as a connecting word linking two noun phrases or clauses. (See Examples TE6-3 and TE6-4.) There are 11 instances of linking 是 (shì) in the collocation 是 你 in the TESOL data (0.31 times per 1000 words), 3 instances of linking 是 (shì) in the LIVT data (0.08 times per 1000 words) and 4 instances of linking 是 (shì) in the NTESOL data (0.09 times per 1000 words.) (See Table 6.54.) We can see that the linking 是 (shì) was heavily used by the TESOL participants in the use of collocation 是 你 [BE you] in comparison with the LIVT and NTESOL participants. The findings suggest that the TESOL

[13] The TESOL data has a noun 问题 ('problem') used in the L1 position in the pattern of 是你. When we checked the instances of 问题 是 你 (= The problem is that you …), we found that there are only two occurrences of this collocation, and these two instances were contributed by the same participant, TE11. The use of this collocation is not very frequent, so we will not discuss it further.

participants tended to make use of the pattern *noun phrase* + 是(shì) + *noun phrase / clause* more often than the LIVT and NTESOL participants in connection with the use of 是 你 [BE you].

TE6-3	问 题 **是** 你 说 真 的 能 够 了 解 什 么 经 济 啊 什 么 吗 ?
	Wèntí **shì** nǐ shuō zhēnde nénggòu liáojiě shénme jīngjì ā shénme ma ?
	The problem **is** that whether you can really understand anything about economics?
TE6-4	它 看 的 **是** 你 的 表 现 , 跟 你 的 反 应 。
	Tā kàn de **shì** nǐ de biǎoxiàn , gēn nǐ de fǎnyìng。
	What it [the grade system] values **are** your performance and your reaction.

	LIVT			TESOL			NTESOL		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是你 (shì nǐ)	17	100	0.44	25	100	0.71	9	100	0.21
Suffix 是(shì) (Adv./Connective+是 (shì))	13	76.47	0.34	14	56.00	0.40	4	44.44	0.09
Linking 是 (shì) (NP+是(shì)+NP/NP+是(shì)+Clause)	3	17.65	**0.08**	11	44.00	**0.31**	4	44.44	**0.09**
Emphasis 是 (shì)	1	5.88	0.03	0	0.00	0.00	1	11.11	0.02

Table 6.54 Use of 是 (shì) in 是你 (shì nǐ) in LIVT, TESOL and NTESOL

The finding that the TESOL participants tended to use the word 是 as a linking word between two noun phrases/clauses matches our finding with regard to the use of *is that* in 6.2.2. The TESOL participants used the collocation *is that* significantly more often than the NTESOL participants, and the collocation *is that* was also used to connect two noun phrases/clauses frequently. The TESOL participants showed a consistency in the frequent use of linking 是 and the linking collocation *is that* in both Mandarin and English. The use of linking 是 with noun phrases/clauses and the linking of clauses by means of *is that* were less apparent in the language of the other groups. This shows that hypothesis 4 is incorrect. What we have found also suggests that the TESOL participants were likely to have been primed differently in the use of the words *is* and 是 in comparison to the other groups of participants (especially the NTESOL participants.) In our discussion of the frequent use of *is that* by the TESOL participants, we argued that the academic and written style of English input might possibly be a factor that had influenced how the TESOL participants used their English. The use of similar patterns in both English and Mandarin (involving the similar lexical items in Mandarin (是) and English (is)) suggests that cross-language influence on

primings of language may exist. There is a possibility that their primings for English use had had an influence on how they used Mandarin and vice versa.

是我 (shì wǒ) [BE I]

The LIVT participants showed a tendency to use the collocation 是我[BE I] more frequently than the TESOL and NTESOL participants. There are 31 instances of 是我 [BE I] found in the LIVT data (0.80 times per 1000 words), 23 instances of 是我 [BE I] found in the TESOL data (0.65 times per 1000 words) and 21 instances of 是我 [BE I] found in the NTESOL data (0.50 times per 1000 words.) (See Table 6.55 and Table 6.56.)

是我 (shì wǒ)	LLR	Sig.		
LIVT - TE	9.511	**0.002**	**	+
LIVT - NTE	14.363	**0.000**	***	+
TE - NTE	0.298	0.585		+

Table 6.55 Log-Likelihood Ratios: 是我 (shì wǒ)

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是 (shì)	1127	100	29.22	941	100	26.65	1139	100	27.04
是我 (shì wǒ)	31	2.75	**0.80**	23	2.44	**0.65**	21	1.84	**0.50**
Patterns of 是我 (shì wǒ)	N L1 Centre R1 1 应 该 是 我 自己 yīnggāi shì wǒ zìjǐ should SHI+I self			N L1 Centre R1 1 问 题 是 我 觉 得 wèntí shì wǒ juédé problem SHI+I think 2 想 要 xiǎngyào want			N L1 Centre R1 1 如果 是 我 觉 得 rúguǒ shì wǒ juédé if SHI+I think 2 自己 zìjǐ self		

Table 6.56 Use of 是 (shì) and 是我 (shì wǒ) in LIVT, TESOL and NTESOL

The patterns of 是我 [BE I] in Table 6.56 show that the use of suffix 是 (shì) and linking 是 (shì) is found in all the groups. When checking the instances of the use of suffix 是 (shì), linking 是 (shì) and emphasis 是 (shì) with the collocation 是我 [BE I], if is found that the LIVT participants used the suffix 是 (shì) more often than the TESOL and NTESOL participants. (See Table 6.57.)

	LIVT		TESOL		NTESOL	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
是我 (shì wǒ)	31	0.80	23	0.65	21	0.50
Suffix 是 (shì) (Adv./Connective+是 (shì))	10	**0.26**	2	**0.06**	5	**0.12**
Linking 是 (shì) (NP+是(shì)+NP/NP+是(shì)+Clause)	19	0.49	20	0.57	13	0.31
Emphasis 是 (shì)	2	0.05	1	0.03	3	0.07
Linking 是(shì) (NP+是(shì)+NP/NP+是(shì)+Clause)	19	0.49	20	0.57	13	0.31
Using Complex noun phrases	1	0.03	1	0.03	0	0

Table 6.57 Use of 是 (shì) in 是我 (shì wǒ) in LIVT, TESOL and NTESOL

LIVT6-2	在 准 备 那 段 期间 **是 我** 觉得 好玩 的 地方 。
	Zài zhǔnbèi nà duàn qījiān **shì wǒ** juédé hǎowán de dìfɑng 。
	What **I** think as the fun part [of the event] **is** when [we were] preparing [for it].
TE6-5	[书] 里面 有 讲 一 句 话 **是 我** 觉得 很 有 , 很 有 [意思 的]。
	[Shū] lǐmiàn yǒu jiǎng yī jù huà **shì wǒ** juédé hěn yǒu , hěn yǒu [yìsi de]。
	There is one sentence in [the book] that **I** think **is** very meaningful.

There are 10 occurrences of suffix 是 (shì) in the use of 是我 [BE I] found in the LIVT data (0.26 times per 1000 words), only 2 occurrences of suffix 是 (shì) in the TESOL data (0.06 times per 1000 words) and 5 occurrences of suffix 是 (shì) in the NTESOL data (0.12 times per 1000 words.) The TESOL participants used the linking 是 (shì) in the collocation 是我 [BE I] more often than the NTESOL participants. There are 20 occurrences of linking 是 (shì) found in the TESOL data (0.57 times per 1000 words), and 13 occurrences of linking 是 (shì) in the NTESOL data (0.31 times per 1000 words). However, when checking the log-likelihood ratios, it is found that the differences between groups are not very great. The LIVT participants used the suffix 是 (shì) in the

collocation 是我 [BE I] more often than the TESOL participants but with weak significance. The other differences are also not significant. There are no great differences between the LIVT and TESOL groups in terms of the frequency of linking 是 (shì) in 是我 [BE I]. (Linking 是 (shì) occurs 0.49 times per 1000 words in the LIVT data.) When checking the surrounding collocates of linking 是 (shì) in the use of 是我 [BE I], no noticeable differences are found. In the use of 是我 [BE I], there are few complex noun phrases involved. There is only one instance found in the LIVT data and one found in the TESOL data. There is no instance of complex noun phrase with 是我 [BE I] found in the NTESOL data. (See Example LIVT6-2 and Example TE6-5.)

Previously when we looked at the use of 是 你 [BE you], we found that the TESOL participants used the linking 是 (shì) in the pattern *noun phrase/clause* + 是 你 + *clause* more frequently than the other participants. Here in the use of 是我 [BE I], the linking 是(shì) was used with similar frequencies by the three groups. (The LIVT and TESOL participants showed greater use of linking 是 (shì) than the NTESOL participants but it did not reach significance.) The significant differences in the use of 是我 [BE I] between the LIVT participants and the TESOL and NTESOL participants are that the LIVT participants made greater use of the suffix 是 (shì) with 我(wǒ) [I] than the TESOL and NTESOL participants. As mentioned, the suffix 是 (shì) is used as part of a connective/adverb. The higher use of suffix 是 (shì) with 我 (wǒ) [I] indicates that the LIVT participants used connectives/adverbs containing 是 (shì) in their Mandarin more often than the TESOL and NTESOL participants when the subject of the sentence is 我 (wǒ) [I]. With regard to the use of collocation 是我 [BE I], the LIVT participants used 是(shì) for discourse purposes more than the other participants. It is hard to tell whether the LIVT participants' English input is the factor that influenced the LIVT participants to make different choices in Mandarin from the others. Nevertheless, the finding of this different discourse preference again shows that our hypothesis 3 is incorrect. Next, we will investigate the use of 是 说 [BE say].

是 说(shì shuō)

The LIVT participants made more use of the collocation 是说 [BE say] than the TESOL and NTESOL participants. The difference in the occurrences of 是说 [BE say] between the LIVT and NTESOL groups is statistically significant. There are 35 occurrences of 是说 in the LIVT data (0.91 times per 1000 words), 22 occurrences of 是说 in the TESOL data (0.62 times per 1000 words) and 20 occurrences of 是说 in the NTESOL data (0.47 times per 1000 words.) The LIVT participants used the collocation 是说 about twice as often as the NTESOL participants. (See Table 6.58 and Table 6.59.)

是说 (shì shuō)	LLR	Sig.		
LIVT - TE	2.782	0.095		+
LIVT - NTE	6.945	**0.008**	****	**+**
TE - NTE	0.763	0.382		+

Table 6.58 Log-Likelihood Ratios: 是 说 (shì shuō) [BE say]

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是(shì)	1127	100	29.22	941	100	26.65	1139	100	27.04
是 说 (shì shuō)	35	3.11	**0.91**	22	2.34	**0.62**	20	1.76	**0.47**

Patterns of 是 说(shì shuō)

N	L1	Centre	R1
1	应 该	是 说	什 麽
	yīnggāi	shì shuō	shénme
	should	SHI+say	what
2	或 者		如果
	huòzhě		rúguǒ
	or/perhaps		if
3			作 文
			zuòwén
			composition

N	L1	Centre	R1
1	或 者	是 说	就 是
	huòzhě	shì shuō	jiùshì
	or/perhaps	SHI+say	just
2	应 该		
	yīnggāi		
	should		

N	L1	Centre	R1
1	或 者	是 说	
	huòzhě	shì shuō	
	or/perhaps	SHI+say	

Table 6.59 Use of 是 (shì) and 是 说 (shì shuō) in LIVT, TESOL and NTESOL

In Table 6.59, we can see that the three groups had similar patterns of the collocation 是说 [BE say]. The collocation 或者 是说 [perhaps/or BE say] was used frequently by all the participants. When examining all the instances of the collocation 是说[BE say], it is found that there are seven colligates (grammatical collocates) found in the L1 position. Before we discuss these, there is a need to say something further about the words 是 (shì) [BE] and 说 (shuō) [to say] briefly.

The word 是 (shì), as already noted, can be used as (1) a suffix to adverbs/connectives (and also as a suffix to the modal verb 应该 [should/ought to]) (2) a linking word connecting two noun phrases or clauses and (3) a marker of emphasis. The word 说 (shuō) is usually (1) a verb meaning 'to say/speak'. However, like the word 是 (shì), the word 说 (shuō) is also used as (2) a suffix to adverbs/connectives. It can also, again like 是 (shì), be used (3) for emphasis. The words 是 (shì) and 说 (shuō), when they are used as a suffix to an adverb or connective, are not interchangeable but they can be used at the same time. When they are used for emphasis, they can be omitted without affecting the intelligibility of the sentences. We can find all of the above use in the language use of all the groups.

There are seven kinds of patterns found with the collocation 是说, which involve: person pronouns, modal verbs, adverbs, connectives, nouns and negative markers. The first kind of patterns is [person pronouns + 是说,]. As Example NTE 6-4 shows, the word 说 (shuō) functions as a verb 'to say'. The word 是 (shì) in this example adds a little emphasis to the action of saying. It can be omitted but the sense of emphasis is lost (see [b]).

NTE6-4	[a] # 如果 让 我 选, 你 **是 说** 未来 工作 吗 ?
	[a] # rúguǒ ràng wǒ xuǎn , nǐ **shì shuō** wèilái gōngzuò ma ?
	If let me choose [it], what you say/mean is 'the job for the future'?
	[b] # 如果 让 我 选 , 你 **说** 未来 工作 吗 ?
	[b] # rúguǒ ràng wǒ xuǎn , nǐ **shuō** wèilái gōngzuò ma ?
	If let me choose [it], you say/mean 'the job for the future'?

The second kind of patterns is [modal verbs + 是说]. As Example LIVT 6-3 shows, the word 说 (shuō) functions as a verb 'to say'. The word 是 (shì) in Example LIVT6-3 [a] adds a little emphasis on the action of saying. The word 是 (shì) can again be omitted without losing meaning [b].

However, if the word 说(shuō) is omitted, the sentence [c] will have a different meaning from that of sentences [a] and [b].

LIVT6-3	[a] 应 该 **是 说** 他 叙述 历史 的 方式 让 我 觉得 [很好]。
	[a] yīnggāi **shì shuō** tā xùshù lìshǐ de fāngshì ràng wǒ juédé [hěnhǎo]。
	[What I] should say is that the way he illustrated history made me feel [good].
	[b] 应 该 **说** 他 叙述 历史 的 方式 让 我 觉得 [很好]。
	[b] yīnggāi **shuō** tā xùshù lìshǐ de fāngshì ràng wǒ juédé [hěnhǎo]。
	[What I] should say is that the way he illustrated history made me feel [good].
	[c] 应该 **是** 他 叙述 历史 的 方式 让 我 觉得 [很好]。
	[c] yīnggāi **shì** tā xùshù lìshǐ de fāngshì ràng wǒ juédé [hěnhǎo]。
	[It] should be that the way he illustrated history made me feel [good].

The third kind of patterns is [adverb + 是说(shìshuō)] and the fourth kind of patterns is [connective + 是说(shìshuō)]. (They function similarly, so we put them together for the purpose of our discussion.) As Example LIVT 6-4 shows, the word 说 (shuō) functions as a marker of emphasis for the action of 'saying'. The word 是 (shì) in Example LIVT6-4 [a] also adds a little emphasis to the action of saying. The word 是 (shì) can be omitted without losing meaning [b]. The word 说 (shuō) can also be omitted in the sentence [c]. Doing so will lose the sense of emphasis, but the sentence meaning will not be too different from that of sentences [a] and [b].

LIVT6-4	[a] [老师] 要求 你 **说** 要 写出 什么 样 的 东西 或者 **是 说** , 你 用 什么 样 的 方法 会 比较 好
	[a] [lǎoshī] yāoqiú nǐ shuō yào xiěchū shénme yàng de dōngxi huòzhě **shì shuō** , nǐ yòng shénme yàng de fāngfǎ huì bǐjiào hǎo
	[The teacher] asked you **(saying that)** write something, or perhaps **(saying that)** it's better for you to use a certain method.
	[b] [老师] 要求 你 **说** 要 写出 什么 样 的 东西 或者 **说** , 你 用 什么 样 的 方法 会 比较 好
	[b] [lǎoshī] yāoqiú nǐ shuō yào xiěchū shénme yàng de dōngxi huòzhě **shuō** , nǐ yòng shénme yàng de fāngfǎ huì bǐjiào hǎo
	[The teacher] asked you **(saying that)** write something, or perhaps **(saying that)** it's better for you to use a certain method.
	[c] [老师] 要求 你 **说** 要 写出 什么 样 的 东西 或者**是** , 你 用 什么 样 的 方法 会 比较 好
	[c] [lǎoshī] yāoqiú nǐ shuō yào xiěchū shénme yàng de dōngxi huòzhě **shì** , nǐ yòng shénme yàng de fāngfǎ huì bǐjiào hǎo
	[The teacher] asked you **(saying that)** write something, or perhaps it's better for you to use a certain method.

The fifth kind of patterns is [negative marker + 是说(shìshuō)]. As Example TE6-6 shows, in sentence [a], the word 说(shuō) functions as a verb 'to say', and the word 是 (shì) functions as linking. With the negative marker involves, the omission of either 是 (shì) or 说 (shuō) will result in a different meaning. (See sentences [b] and [c] in Example TE6-6.)

TE6-6	[a] [我] 不 **是 說** [我] 不 喜歡 寫 文章。
	[a] [wǒ] bú **shì shuō** [wǒ] bú xǐhuɑn xiě wénzhāng。
	What [I] am saying is not that [I] don't like writing an essay.
	[b] [我] 不 **是** 不 喜歡 寫 文章。
	[b] [wǒ] bú **shì** bú xǐhuɑn xiě wénzhāng。
	It is not that I don't like writing an essay. (It's something else.)
	[c] [我] 不 **說** [我] 不 喜歡 寫 文章。
	[c] [wǒ] bú **shuō** [wǒ] bú xǐhuɑn xiě wénzhāng。
	[I] won't say that I don't like writing an essay.

The sixth kinds of pattern is [noun/noun phrase + 是说(shìshuō)]. As Example TE6-7 shows, in sentence [a] the word 说 (shuō) functions as a marker of emphasis on 'saying', and the word 是(shì) functions as linking. The omission of 说 (shuō) is acceptable but the omission of 是 (shì) will be unnatural to a Mandarin speaker. (See sentences [b] and [c].)

TE6-7	[a] 你 的 意思 **是 說** , 很 小 那 種 小朋友 嗎 ?
	[a] Nǐ de yìsi **shì shuō** , hěn xiǎo nà zhǒng xiǎopéngyou mɑ ?
	What you mean in your **words (saying) is** that kind of very young children?
	[b] 你 的 意思 **是** , 很 小 那 種 小朋友 嗎 ?
	[b] Nǐ de yìsi **shì** hěn xiǎo nà zhǒng xiǎopéngyou mɑ ?
	What you mean **is** that kind of very young children?
	[c] *****你 的 意思 **說** , 很 小 那 種 小朋友 嗎 ?
	[c] *****Nǐ de yìsi **shuō** , hěn xiǎo nà zhǒng xiǎopéngyou mɑ ?

The seventh and last kind of patterns is [verb/verb phrase + 是说(shìshuō)]. In Example NTE6-5, the word 说 (shuō) functions a little differently from the previous ones. Although it still adds a sense of action 'saying' in the sentence, it is more similar to a pause filler. If the word 说 (shuō) is taken away, it will not affect the meaning of the sentences. The word 是 (shì) functions as linking. However, the omission of 是 (shì) is acceptable due to the transitive verb in use. (See sentence [d].)

NTE6-5	[a] [古文] 就 變 成 **是 說** 你 可以 就是 直接 看 [它]。
	[a] [gǔwén] jiù biànchéng **shì shuō** nǐ kéyǐ jiùshì zhíjiē kàn [tā]。
	The Classics would transform to be, **say**, [something that] you can read directly.
	[b] [古文] 就 變 成 **說** 你 可以 就是 直接 看 [它]。
	[b] [gǔwén] jiù biànchéng **shuō** nǐ kéyǐ jiùshì zhíjiē kàn [tā]。
	The Classics would transform to, **say,** [something that] you can read directly.
	[c] [古文] 就 變 成 **是** 你 可以 就是 直接 看 [它]。
	[c] [gǔwén] jiù biànchéng **shì** nǐ kéyǐ jiùshì zhíjiē kàn [tā]。
	The Classics would transform to be (something that) you can read directly.
	[d] [古文] 就 變 成 你 可以 就是 直接 看 [它]。
	[c] [gǔwén] jiù biànchéng nǐ kéyǐ jiùshì zhíjiē kàn [tā]。
	The Classics would transform to (something that) you can read directly.

When checking all the instances of 是说 used by the three groups, it is found that the LIVT participants used the patterns [Person pronoun + 是说] and [Modal verb + 是说] more often than the participants in the other groups. The LIVT participants used the pattern [person pronoun + 是说] 0.29 times per 1000 words. The TESOL participants used this pattern 0.11 times per 1000 words, and the NTESOL participants used this pattern 0.19 times per 1000 words. The LIVT participants used the pattern [person pronoun + 是说] twice as often as the TESOL participants. The LIVT participants used the pattern [Modal verb + 是说] 0.21 times per 1000 words. The TESOL participants used this pattern 0.06 times per 1000 words, and the NTESOL participants used it only 0.02 times per 1000 words. The LIVT participants used the pattern [Modal verb + 是说] more than three times as often as the TESOL participants, and 10 times as often as the NTESOL participants. (See Table 6.60.) When checking the log-likelihood ratios, the LIVT participants used the pattern [Modal verb + 是说] significantly more frequently than the NTESOL participants.

The two patterns we have been considering use the word 是 (shì) as a form of emphasis. The above findings suggest that the LIVT participants used the word 是 (shì) for discourse purposes more often than the other two groups of participants when using 是说, especially more often than the NTESOL participants. In fact, when we reviewed our earlier investigation of the use of 我是 [I BE] and 是我 [BE I] in the LIVT group, it can be found that the LIVT participants also used 是 (shì) for discourse purposes more often than the NTESOL participants. Although we cannot know exactly why the LIVT participants used the word 是 (shì) in its discourse function more often than the NTESOL group, these findings show that our hypothesis 3 is incorrect. There are noticeable differences in the use of spoken Mandarin between the participants studying in the UK and the ones in Taiwan.

	LIVT			TE			NTE		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
是 说 (shì shuō)	35	100	0.91	22	100	0.62	20	100	0.47
Person pronoun +是 说 (shì shuō)	11	31.43	**0.29**	4	18.18	**0.11**	8	40.00	**0.19**
Modal verb+是 说 (shì shuō)	8	22.86	**0.21**	2	9.09	**0.06**	1	5.00	**0.02**
Adverb/Connective+是 说 (shì shuō)	5	14.29	0.13	3	13.64	0.08	3	15.00	0.07
Negative M. +是 说 (shì shuō)	7	20.00	0.18	8	36.36	0.23	6	30.00	0.14
Noun/NP+是 说 (shì shuō)	2	5.71	**0.05**	4	18.18	**0.11**	1	5.00	**0.02**
Verb+是 说 (shì shuō)	0	0.00	0.00	0	0.00	0.00	1	5.00	0.02
Unable to be placed[14]	2	18.18	0.05	1	25.00	0.03	0	0.00	0.00

Table 6.60 Patterns of 是 说 (shì shuō)[BE say] in LIVT, TESOL and NTESOL

In Table 6.60, we can also see that the TESOL participants used the pattern [Noun/noun phrase + 是说[shìshuō]] more often than the participants in the other two groups. The use of 是 (shì) in this pattern is a linking word connecting noun phrases or clauses. The TESOL participants used the pattern 0.11 times per 1000 words. The LIVT participants used this pattern 0.05 times per 1000 words, and the NTESOL participants only used it 0.02 times per 1000 words. The TESOL participants used the pattern [Noun/noun phrase + 是说[shìshuō]] twice as often as the LIVT participants, and 5 times as often as the NTESOL participants. Although the differences are not significant in the log-likelihood ratios, taken in connection with our previous investigation into the use of 是你[shìnǐ][BE you], the findings suggest that the TESOL participants used the word 是[shì] [BE] in its linking function more often in their spoken Mandarin in comparison with the other participants. This special feature in the use of 是[shì] [BE] in the TESOL group further confirms that our hypothesis 4 is incorrect.

[14] The use of 是说[shìshuō][BE say] in these instances were in uncompleted sentences, so it is impossible to categorise them with certainty.

6.3.3 Summary of the use of *是 (shì)*

In our investigation on the use of 是 (shì), we observed how the three groups of participants used the collocations 我是 [I BE], 它是 [it BE], 是你 [BE you], 是我 [BE I] and 是说[BE say]. The findings are presented in Table 6.61 and Table 6.62.

In our investigation on these collocations, we found that all three groups shared many of the same collocations for these collocations. This finding is potentially compatible with the concept of **nesting** in lexical priming theory. (Hoey, 2005: 8) The shared collocations (or called word sequences) containing these words become a unit loaded with the contexts and co-texts, and then prime the individual as a whole unit instead of the separate words. The LIVT and TESOL participants were found to use the collocation 我是 [I BE] more often than the NTESOL participants. We also found that the LIVT and TESOL participants used the word 是 (shì) for adding emphasis to what they were doing or to what they were relating more often than the NTESOL participants. The LIVT and TESOL participants also did not use 它是[it BE] in the way that the NTESOL participants used it in the collocation 它是[it BE.] With regard to the use of 是说 [BE say], the LIVT participants used the patterns [Person pronoun + 是说] and [Modal verb + 是说] more often than the other two groups of participants. We also found that the LIVT participants used the word 是(BE) for discourse purposes (emphasis) more often than the other two groups in the contexts of the use of 是说[BE say]. These findings indicate that although participants used similar collocations of a word/word sequence, they were still behaving differently in other respects. These differences challenge our third and fourth hypotheses in which we assumed that the three groups of participants would not show any sign of differences in their Mandarin.

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	Collocations	The types of differences	The possible factors that affected this difference OR noteworthy observation
The LIVT participants used the collocations **significantly more frequently** than the NTESOL participants.	我是 (wǒ shì) [I BE]	Discourse Preference	The LIVT participants were prone to add emphasis to the verbs with the word 是 (shì) when the subjects referred to themselves more often than the NTESOL participants did.
	是你 (shì nǐ) [BE you]	Discourse Preference	The LIVT participants were prone to use the suffix 是 (shì) (as part of an adverb/a connective) in the use of 是你 (shìnǐ) [BE you] more frequently than the NTESOL participants.
The LIVT participants used the collocations **significantly less frequently** than the NTESOL participants.	它是 (tā shì) [it BE]	Discourse Preference	The LIVT participants were not prone to use the pronoun 它(tā) [it] when describing a thing's features or qualities, and also when characterising the thing with a noun phrase, as often as the NTESOL participants were. The LIVT participants tended not to add emphasis with the use of 它 (tā) [it] as often as the NTESOL participants. *(The LIVT and TESOL participants possibly had different primings for the use of 它 (tā) [it] from the NTESOL participants.)*
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	是我 (shì wǒ) [BE I]	Discourse Preference	The LIVT participants were prone to use the suffix 是 (shì) (as part of an adverb/a connective) with 我 (wǒ) [I] more often than the TESOL and NTESOL participants.
	是说 (shì shuō) [BE say]	Discourse Preference	The LIVT participants were prone to use the patterns [Person pronoun + 是 (shì) 说 (shuō)] and [Modal verb + 是 (shì) 说 (shuō)] more often than the other two groups of participants. (These two patterns use the word 是 (shì) as a form of emphasis.)

Table 6.61 Findings in the use of 是 (shì) that disagree with Hypothesis 3

We found that the TESOL participants tended to use the pattern [noun phrase + 是 (shì)+ noun phrase / clause] more often than the LIVT and NTESOL participants in the contexts of use of 是你. The TESOL participants also used the pattern [Noun/noun phrase+是说] more often than the other two groups. The use of 是 (shì) in this pattern is as a linking word connecting noun phrases or clauses. The TESOL participants were more likely to use linking 是 [BE] in their spoken Mandarin. Although these kinds of use were found in all the groups, the fact that the differences of language use which show statistical significance between the LIVT and TESOL participants and the NTESOL

participants suggests that there are some factors other than their Mandarin primings influencing the LIVT and TESOL participants' Mandarin use.

Findings that Disagree with Hypothesis 4 (null hypothesis)			
Noticeable Differences	Collocations	The types of differences	The possible factors that affected this difference OR noteworthy observation
The TESOL participants used collocations **significantly more frequently** than the NTESOL participants.	我是 (wǒ shì) [I BE]	Discourse Preference	The TESOL participants were prone to add emphasis with the word 是 (shì) on the verbs more often than the NTESOL participants, when the subjects of the clause referred to themselves.
	是你 (shì nǐ) [BE you]	Discourse Preference/ Lexico-grammatical Difference	The TESOL participants were prone to use the suffix 是 (shì) (as part of an adverb/a connective) in the use of 是你 (shì nǐ) [BE you] more frequently than the NTESOL participants. (Discourse Preference) The TESOL participants made use of the pattern [noun phrase + 是(shì) + noun phrase / clause] more often than the LIVT and NTESOL participants in connection with the use of 是你 (shì nǐ) [BE you].(Lexico-grammatical Difference)
The TESOL participants used the collocations **significantly less frequently** than the NTESOL participants.	它是 (tā shì) [it BE]	Discourse Preference	The TESOL participants did not use the pronoun 它 (tā) [it] as often as the NTESOL participants when describing something's features or qualities. When characterising the things with a noun phrase, the TESOL participants did not add emphasis with the use of 它 (tā) [it] as often as the NTESOL participants.

Table 6.62 Findings in the use of 是 (shì) that Disagree with Hypothesis 4

6.4 Conclusion

In this chapter, we have investigated the use of the English word *is/'s* and the Mandarin word 是 [BE] in order to test our four hypotheses. In our investigation on the use of *is* in the three groups, the first and the second hypotheses were partly supported. The LIVT and the TESOL participants did show noticeable differences from the NTESOL participants, and the different use of English is very likely to had been shaped by the influence from their English input. In our investigation on the use of 是(shì) [BE] in the Mandarin of the three groups, we found that our null hypotheses (the third and fourth hypotheses) were challenged. Although there were great similarities in the use of collocations of the Mandarin word 是 (shì), we did find several noticeable differences between the participants in the three groups. The third and fourth hypotheses were therefore incorrect. A summary of the findings of the use of *is* is presented in Table 6.63 and Table 6.64, and a summary of the findings of the use of 是 (shì) [BE] is presented in Table 6.65 and Table 6.66.

Summary of the findings that Agree with Hypothesis 1 – Discourse Preference	
Words	Descriptions OR noteworthy findings/observation
is quite/'s quite	The LIVT participants were prone to use the adverb *quite* to modify adjectives significantly more often than the TESOL and NTESOL participants. There were **more** participants in the LIVT group using *is ('s) quite* in comparison with the ones in the other two groups.
Summary of the findings that Agree with Hypothesis 1 – Lexico-grammatical Difference	
Words	Descriptions OR noteworthy findings/observation
that is/that's	The collocates of *that is ('s)* used by the LIVT participants are similar to the ones found in the reference corpora, BASEah and COCAsp. The NTESOL participants did not share this similarity.

Table 6.63 Summary of the findings of the use of *is* that Agree with Hypothesis 1

Summary of the findings that Agree with Hypothesis 2 – Discourse Preference	
Words	Descriptions OR noteworthy findings/observation
is pretty/'s pretty	The TESOL participants were prone to use the adverb *pretty* to modify adjectives significantly more often than the LIVT and NTESOL participants. There were more participants in the TESOL group using *is ('s) pretty* in comparison with those in the other two groups.
it's	The use of *I think it is/it's because* was used as a relatively fixed expression when the LIVT and NTESOL participants used the collocation *it is/it's because*. The TESOL participants did not share this tendency.
Summary of the findings that Agree with Hypothesis 2 – Lexico-grammatical Difference	
Words	Descriptions OR noteworthy findings/observation
that is/that's	The collocates of *that is ('s)* used by the TESOL participants are similar to the ones found in the reference corpora, BASEah and COCAsp. The NTESOL participants did not share this similarity.
is that	The TESOL participants were prone to use *is that* in the pattern [*NP+ is that + [S]*] significantly more often than the LIVT and NTESOL participants.

Table 6.64 Summary of the findings of the use of *is* that Agree with Hypothesis 2

Our findings suggest that the hypotheses 1 and 2 are correct. There are noticeable differences in language use between the participants studying in the UK and the participants studying in Taiwan, and between the participants studying in English-relevant subjects and those studying in non-English-relevant subjects. The differences that we found are of two kinds. The first kind is a difference of discourse preferences. One group of participants showed higher or lower frequency of the use of words/word sequences for particular discourse purposes (e.g. emphasis or modality) than the other groups of participants. The second kind of differences is lexico-grammatical difference. One group of participants showed higher or lower frequency of the use of words in particular lexico-grammatical patterns/collocations than the other groups of participants. With regard to differences in discourse preferences, we found that the LIVT participants used the adverb *quite* (*is / ('s) quite)* to modify their adjectives more often than the other participants, and the TESOL participants, likewise, used the adverb *pretty* (*is / ('s) pretty*) to modify their adjectives more often than the other participants. With regard to lexico-grammatical differences, we found the LIVT and TESOL participants had different patterns of collocates in the use of *that is (that's)* from the NTESOL participants, and the LIVT and TESOL participants' use of the collocates of *that is(that's)* was similar

to those used by the speakers in the BASEah and COCAsp corpora. Although we cannot make a firm claim that the different collocations in the use of *that is (that's)* in the LIVT and TESOL groups is a result of the English input that these two groups of participants had received, since we cannot access to the exact English input they had received, the findings of the collocations in the use of *that is(that's)* by the LIVT and TESOL participants suggests that the cause of this difference is very likely relevant to their different English input/learning experience. A similar situation can be seen in the findings that TESOL participants used *is that* in the pattern [*NP+ is that + [S]*] more frequently than the LIVT and NTESOL participants. Similarly, we cannot firmly say the difference is caused by their English input or educational training, because we have not had sufficient access to their English input and the content of their training. However, the significant differences between the TESOL participants and the participants in the other groups in this respect show that different input/training may well be the possible explanation to the findings.

Summary of the findings that disagree with Hypothesis 3 – Discourse Preference	
Words	Descriptions OR noteworthy findings/observation
我是 (wǒ shì) [I BE]	The LIVT participants were prone to add emphasis to the verbs that follow with the word 是 (shì) more often than the NTESOL participants did, when the subjects of the clauses referred to themselves
是你 (shì nǐ) [BE you]	The LIVT participants were prone to use the suffix 是 (shì) (as part of an adverb/a connective) in the context of 是你(shì nǐ) [BE you] more frequently than the NTESOL participants.
它是 (tā shì) [it BE]	The LIVT participants did not use the pronoun 它 (tā) [it] as often as the NTESOL participants when describing something's features or qualities, When characterising the things with a noun phrase, the LIVT participants were not prone to add emphasis with the use of 它(tā) [it] as often as were the NTESOL participants.
是我 (shì wǒ) [BE I]	The LIVT participants used the suffix 是 (shì) (as part of an adverb/a connective) with 我 (wǒ) [I] more often than the TESOL and NTESOL participants.
是说 (shì shuō) [BE say]	The LIVT participants were prone to use the patterns [*Person pronoun + 是(shì)说(shuō)*] and [*Modal verb + 是(shì)说(shuō)*] more often than the other two groups of participants. (The word 是 (shì) is used as a form of emphasis in these two patterns.)
Summary of the findings that disagree with Hypothesis 3 – Lexico-grammatical Difference	
Words	Descriptions OR noteworthy findings/observation
N/A	N/A

Table 6.65 Summary of the findings that Disagree with Hypothesis 3

Summary of the findings that Disagree with Hypothesis 4 – Discourse Preference	
Words	Descriptions OR noteworthy findings/observation
我是 (wǒ shì) [I BE]	The TESOL participants were prone to add emphasis to the verbs that follow with the word 是 (shì) more often than the NTESOL participants when the subjects of their clauses were themselves
是你 (shì nǐ) [BE you]	The TESOL participants used the suffix 是 (shì) (as part of an adverb/a connective) in connection with the use of 是你 (shì nǐ) [BE you] more frequently than the NTESOL participants.
它是 (tā shì) [it BE]	The TESOL participants were not prone to use the pronoun 它(tā) [it] when describing something's features or qualities. When characterising the things with a noun phrase, the TESOL participants were not prone to use the pronoun 它(tā) [it] as often as the NTESOL participants. The TESOL participants did not add emphasis with the use of 它(tā) [it] as often as the NTESOL participants.
Summary of the findings that disagree with Hypothesis 4 – Lexico-grammatical Difference	
Words	Descriptions OR noteworthy finding/observation
是你 (shì nǐ) [BE you]	The TESOL participants made use of the pattern [*noun phrase + 是(shì) + noun phrase / clause*] more often than the LIVT and NTESOL participants in connection with the use of 是你(shì nǐ) [BE you].

Table 6.66 Summary of the findings that disagree with Hypothesis 3

Our findings suggest that the hypotheses 3 and 4 are incorrect. With regard to the difference in the discourse preferences, we found that the LIVT participants used the word 是(shì) for discourse purposes more often than the other groups of participants, especially more than the NTESOL participants. (See Table 6.65.) The TESOL participants also, although the tendency was not as strong as for the LIVT participants, used 是 (shì) for discourse purposes more often than the NTESOL participants. (See Table 6.66.) These different discourse preferences that the participants had are unexpected, since the participants were asked identical questions in a similar context. However, the causes of these discourse preferences, as we discussed at the beginning of chapter 4, may be social or educational factors. It is unclear whether there is influence from the different English input/experience/training of the three groups. However, with the regard to lexico-grammatical differences, we found that the TESOL participants used the collocation 是 你 [BE you] in the pattern [*noun phrase + 是(shì) + noun phrase / clause*] more frequently than the other groups of participants. The greater use of this pattern matches our finding of the use of *is that* in the TESOL

participants' English use. In these two patterns, the use of the word 是 (shì) and the words *is (that)* serve a similar function of linking noun phrases/clauses. The fact that the TESOL participants showed consistency in using the linking 是 (shì) and the linking word sequence *is that* more frequently than the other groups of participants suggests that their different language input may play a role in their language use.

Most of the time in our observation of frequently used collocations, we did not find great differences in terms of the nature of collocations, colligations or semantic associations, there were statistically significant differences in quantity of usage. These differences in quantity of usage between groups show the possibility that the participants of one group may be primed to use certain collocations/patterns or discourse strategies more strongly/weakly than the participants of the other groups in a similar context. These differences also show the possibility that there were more/fewer participants primed to use certain collocation/patterns or discourse strategies in one particular group than in the other groups. (e.g. with the use of *pretty* in the TESOL group and the use of *quite* in the LIVT group.) However, some of our findings suggest that some Taiwanese participants had likely been influenced in their priming of language use which differentiated them from the rest of the Taiwanese participants as regards the use of certain Mandarin words/word sequences. (e.g. the impersonal pronoun 它 in our discussion of 它是 [it BE] in 6.3.1.)

In our hypotheses, we predicted that the different English input backgrounds would have no effect on these participants' use of spoken Mandarin. As what has been discussed in chapter 2, if the way a person uses a language is built up by the language input to which s/he has been exposed, there should be no great differences in these Taiwanese participants' use of spoken Mandarin. Their Mandarin use should be quite similar to each other, especially as regards the lexico-grammatical use of words. These participants were primed to use these kinds of words from the first day they were exposed to Mandarin, and one would therefore have predicted that their use of lexico-grammatical words/word sequences would be unlikely to differ greatly. Of course we did find that such use of words/word sequences were mostly the same/very similar across all the groups. However, the findings of the significant differences are intriguing and require for further research.

The factors that caused the participants in the LIVT and TESOL groups to use Mandarin differently from the participants in the NTESOL group are worth studying. We propose that it is likely that there is a cross-language influence on the participants' primings of words and this kind of influences in two ways. A person's priming of the use of native language (in this case, Mandarin) will possibly affect how s/he uses the target language (English) This notion is close to 'first/native language interference' in second language acquisition (Dulay, Burt and Krashen, 1982), which is an automatic language transfer of the surface structure of the native language to the surface of the target

language. What we have found suggests that a person's priming of the use of target language (English) will also possibly influence his/her priming of the use of native language (Mandarin). The influences will be most likely on the lexical items in both languages which share some similarity in their original uses. The fact that the LIVT and TESOL participants were more likely to had been exposed to a greater amount and variety of English input than the NTESOLparticipants may imply that the way a participant in either of these groups may be affected by his/her English primings when using Mandarin. It is recognised in SLA that the first language often affects how a learner uses the target language. However, it is seldom mentioned that a learner's use of first language may be affected by the target language. It is possible that cross-language priming of lexico-grammatical language use exists. However, the current study can only point to the possibility. Much more research on a wider range of items and a larger body of data would be needed to provide convincing proof. Nevertheless, from what we have observed, it appears that this phenomenon may exist. Chapter 7, which looks at the English word *have* and Chinese word 有 (yǒu) [have], will examine more evidence with regard to both the null hypotheses (3 and 4) and the English hypotheses (1 and 2).

Chapter 7 The Use of HAVE and 有 (yǒu) by the three Taiwanese groups

7.1 Introduction

In this chapter, we will investigate the English word *have* and the Chinese word 有(yǒu). The word *have* is the 26th most frequent word in the LIVT group, the 23rd most frequent word in the TESOL group and the 18th most frequent word in the NTESOL group. The Chinese word 有 (yǒu) is in all three top 10 most frequent word lists for the Taiwanese groups. (See Table 7.1 and Table 7.2.)

have	Rank	Freq.	Per 1000 words
LIVT	26	227	7.86
TE	23	264	8.89
NTE	18	250	10.73

Table 7.1 Frequencies of *have* in LIVT, TESOL and NTESOL

有 (yǒu)	Rank	Freq.	Per 1000 words
LIVT	7	870	22.56
TE	5	870	24.64
NTE	5	1059	25.14

Table 7.2 Frequencies of 有 (yǒu) in LIVT, TESOL and NTESOL

We will look at the use of English *have* and Mandarin 有(yǒu) together for our investigation. The reason is similar to that for choosing to investigate *is* and 是(shì) in the previous chapter. The English word *have* is often translated as the Chinese word 有(yǒu) in the Chinese-English dictionary. The word *have* is used to express the possession of certain things, and so is the word 有 (yǒu). (我 wǒ 有 yǒu 台 tái 车 chē = I have a car.) The word *have* is used to express tense, aspect and modality. The Chinese word 有 (yǒu) can be used to express duration and existence of an event/thing, similar to the past perfect use in English. Unlike the word *have*, the Chinese word 有 (yǒu) is often used to form an adjective or an adverb (by combining with a noun or an adjective), such as 有 yǒu 心 xīn [have + heart] which means 'on purpose' or 'thoughtful') or 富有 fù yǒu ([rich + have] which means 'rich'). Despite of the differences between these two words, the similarity of the English *have* and the Chinese 有(yǒu) makes them a reasonable pair for observation.

We continue to test our hypotheses, and the hypotheses are repeated here for convenience. With respect to the use of English:

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

With respect to the use of Mandarin Chinese:

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

In Table 7.3, we can see that the LIVT participants used the word *have* significantly less frequently than the TESOL and NTESOL participants. There are 7.86 instances of the use of *have* per 1000 words in the LIVT data, 8.89 instances per 1000 words in the TESOL data and 10.73 instances per 1000 words in the NTESOL data.

		Freq.	Per 1000 words		LLR	Sig.		
have	LIVT	227	7.86	LIVT-TE	1.856	0.173		-
	TESOL	264	8.89	**LIVT-NTE**	11.492	**0.001**	***	-
	NTESOL	250	10.73	**TE-NTE**	4.499	**0.034**	*	-

Table 7.3 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL

When talking all forms of the use of *HAVE-verbs* (*has, had, 've*) into concern, the differences between the LIVT and TESOL participants disappeared. However, the NTESOL participants showed significantly more use of *HAVE-verbs* in comparison with the LIVT participants. (See Table 7.4)

		Freq.	Per 1000 words		LLR	Sig.		
HAVE-Verbs (VH)	LIVT	261	9.04	LIVT-TE	2.071	0.150		-
	TESOL	303	10.21	**LIVT-NTE**	6.666	**0.010**	**	-
	NTESOL	264	11.33	TE-NTE	1.527	0.217		-

Table 7.4 Frequencies of *HAVE-Verbs* in LIVT, TESOL and NTESOL

There is one noticeable difference when checking the total use of *HAVE-Verbs*. The use of the abbreviated *have* (*'ve*) is significantly more frequently used by the LIVT and TESOL participants than by the NTESOL participants. There are 19 occurrences of the use of *'ve* found in the LIVT data (0.66 times per 1000 words), 10 occurrences of the use of *'ve* found in the TESOL data (0.34 times per 1000 words), but only 2 occurrences of *'ve* found in the NTESOL data (0.09 times per 1000 words). (See Table 7.5.)

HAVE-Verbs (VH)	LIVT			TESOL			NTESOL		
	261	100	9.04	303	100	10.21	264	100	11.33
	Freq.	%	per 1000 words	Freq.	%	per 1000 words	Freq.	%	per 1000 words
have	227	86.97	7.86	264	87.13	8.89	250	94.70	10.73
has	5	1.92	0.17	14	4.62	0.47	6	2.27	0.26
had	10	3.83	0.35	15	4.95	0.51	6	2.27	0.26
_'ve	19	7.28	**0.66**	10	3.30	**0.34**	2	0.76	**0.09**

Table 7.5 Use of *HAVE-Verbs* in LIVT, TESOL and NTESOL

Table 7.6 shows that the use of *'ve* is significantly more frequent in the LIVT and TESOL groups than in the NTESOL group. The use of *'ve* is twice more frequent in the LIVT group than in the TESOL group, and it is seven times more frequent than in the NTESOL group. The use of *'ve* in the TESOL group is almost four times more frequent than in the NTESOL group. This is quite a striking finding if we consider that it is the NTESOL group that showed a higher frequency of the total use of *HAVE-Verbs*.

LIVT - NTE					LIVT - TE			TE - NTE				
Over					Over			Over				
Word	LLR	Sig.			Word	LLR	Sig.	Word	LLR	Sig.		
_'ve	12.503	**0.000**	***	+				**_'ve**	4.061	**0.044**	*	+

Table 7.6 Log-Likelihood Ratios: *'ve* in LIVT, TESOL and NTESOL

The word *have* in its abbreviated form *'ve* is often serving as an auxiliary verb for the perfect aspect. The frequent use of *'ve* in the LIVT and TESOL groups suggests that these two groups of participants

seemed to be prone to employ the word *have* in this auxiliary function, not for its regular process verb function. We will test this assumption in 7.2.2. First, we will investigate the L1 collocates of *have*.

7.2 The use of *have*

7.2.1 L1 Collocates of *have*

Table 7.7 presents the top 10 most frequent L1 collocates of *have* in the language of the three Taiwanese groups. The use of pronouns *I, you, we* and *they* as frequent L1 collocates is found in all the three groups. The words *to* and *don't* are found as frequent L1 collocates in all the three groups. There are several words that appear only in one or two groups' top 10 most frequent lists; for instance, the adverb *still* appears in only the LIVT and NTESOL groups' lists, and the modal verb *will* appears in only the LIVT and TESOL groups' lists. We then examined these collocates via log-likelihood ratios to see whether there are any noticeable differences worth investigating.

	LIVT	227	7.86	TESOL	264	8.89	NTESOL	250	10.73
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	I	63	2.18	I	86	2.90	I	101	4.33
2	YOU	37	1.28	YOU	42	1.41	DON'T	29	1.24
3	DON'T	31	1.07	WE	33	1.11	WE	25	1.07
4	THEY	13	0.45	THEY	24	0.81	YOU	21	0.90
5	TO	12	0.42	DON'T	22	0.74	THEY	19	0.82
6	WILL	10	0.35	WILL	5	0.17	EH	6	0.26
7	WE	7	0.24	TO	4	0.13	IT'S	5	0.21
8	STILL	6	0.21	CAN	4	0.13	TO	4	0.17
9	JUST	5	0.17	ALWAYS	4	0.13	HAVE	4	0.17
10	IT	4	0.14	ONLY	3	0.10	STILL	2	0.09

Table7.7 L1 Collocates of *have* in LIVT, TESOL and NTESOL

In Table 7.8, it is seen that the LIVT and TESOL participants used the collocation *I have* (not in the short form) significantly less frequently than the NTESOL participants. In the previous section, we found that there is a noticeable difference between the LIVT and TESOL participants and the NTESOL participants in the use of the abbreviated form *'ve*. We now note that the use of *I've* was also used significantly more often by the LIVT and TESOL participants than by the NTESOL participants. (See Table 7.9.) In fact, every instance of the use of *'ve* in the LIVT and TESOL data takes the form of *I've*. There are only two instances of the use of *'ve* in the NTESOL data, and one of them takes the form of *I've*. (See Table 7.10.) The distinct difference between the LIVT and TESOL groups and the NTESOL group is intriguing. Therefore, we will focus on the use of *I have*/*I've*.

LIVT - NTE					LIVT - TE					TE - NTE				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Words	LLR	Sig.		
WILL HAVE	6.746	0.009	**	+	TO HAVE	4.412	0.036	*	+					
					STILL HAVE	4.103	0.043	*	+					
Less					Less					Less				
Words	LLR	Sig.			Words	LLR	Sig.			Word	LLR	Sig.		
WE HAVE	14.961	0.000	***	-	WE HAVE	17.638	0.000	***	-	**I HAVE**	7.565	**0.006**	****	**-**
I HAVE	18.899	**0.000**	****	**-**						IT'S HAVE	3.967	0.046	*	-

Table 7.8 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL – L1 Collocates

LIVT - NTE					LIVT - TE					TE - NTE				
Over					Over					Over				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
I've	16.160	**0.000**	****	**+**	I've	3.096	0.078		+	**I've**	6.530	**0.011**	****	**+**

Table 7.9 Log-Likelihood Ratios: *I've* in LIVT, TESOL and NTESOL

* _ 've	LIVT		19	0.66	TESOL		10	0.34	NTESOL		2	0.09
N	L1	With	Freq.	per 1000 words	L1	With	Freq.	per 1000 words	L1	With	Freq.	per 1000 words
1	**I**	**_'VE**	19	**0.66**	**I**	**_'VE**	10	**0.34**	**I**	**_'VE**	1	**0.04**
2									THEY	_'VE	1	0.04

Table 7.10 Use of *'ve* in LIVT, TESOL and NTESOL

In Table 7.8, it is found that the LIVT participants used the collocation *we have* less often than the TESOL and NTESOL participants. The significantly high frequency of the collocation *we have* in the TESOL and NTESOL data suggests that these two groups of participants used the pronoun *we* as a general pronoun to illustrate a general statement more often than the LIVT participants. It is similar to what we have discussed in chapter 4, namely that the LIVT participants tended to use the second person pronoun as the general noun in their speech and the other groups of participants showed no such a tendency. Although this special tendency to use one particular pronoun as the general pronoun is interesting language behaviour, it does not provide us too much information for testing our hypotheses. Therefore, the use of *we have* will not be investigated in detail. There are other collocations *will have*, *still have*, *it's have* and *to have* used by one group of participants significantly more often than by the other groups of participants. However, given that their actual occurrences do not differ greatly between groups, we will not include these collocations in our further investigation. We will however investigate further the use of *I have* and *I've*.

I HAVE / I'VE

The use of *I have* is significantly more frequent in the NTESOL data than in the LIVT and TESOL data. In the NTESOL data, the use of *I have* accounts for 40.40% of the total use of *have* (4.33 occurrences per 1000 words.) The use of *I have* accounts for 32.58% of the total use of *have* (2.90 occurrences per 1000 words) in the TESOL data, and it only accounts for 27.75% of the total use of *have* (2.18 times per 1000 words) in the LIVT data. We can see that the LIVT and TESOL participants did not differ much in terms of the frequencies of the use of *I have*. The NTESOL participants, however, used the collocation *I have* twice as often as the LIVT participants in terms of absolute frequencies and about 1.5 times more often in terms of proportional frequencies. However, the patterns of the collocation *I have* in the three groups did not differ much. The pattern *connectives + I have + to* is the most common pattern in all the groups. (See Table 7.11.)

<table>
<tr><td></td><td colspan="3">LIVT</td><td colspan="3">TE</td><td colspan="3">NTE</td></tr>
<tr><td></td><td>Freq.</td><td>%</td><td>Per 1000 words</td><td>Freq.</td><td>%</td><td>Per 1000 words</td><td>Freq.</td><td>%</td><td>Per 1000 words</td></tr>
<tr><td>HAVE</td><td>227</td><td>100</td><td>7.86</td><td>264</td><td>100</td><td>8.89</td><td>250</td><td>100</td><td>10.73</td></tr>
<tr><td>I HAVE</td><td>63</td><td>27.75</td><td>2.18</td><td>86</td><td>32.58</td><td>2.90</td><td>101</td><td>40.40</td><td><u>4.33</u></td></tr>
<tr><td>I'VE</td><td>19</td><td>--</td><td><u>0.66</u></td><td>10</td><td>--</td><td><u>0.34</u></td><td>1</td><td>--</td><td>0.04</td></tr>
<tr><td>R1 collocates of I have</td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>SO</td><td>I HAVE</td><td>TO</td></tr><tr><td>2</td><td>IF</td><td></td><td>EH</td></tr><tr><td>3</td><td>EH</td><td></td><td>NO</td></tr><tr><td>4</td><td>WHY</td><td></td><td>THE</td></tr><tr><td>5</td><td>THINK</td><td></td><td>OPPORTUNITY</td></tr><tr><td>6</td><td>HAVE</td><td></td><td>ASK</td></tr><tr><td>7</td><td>ASK</td><td></td><td>MORE</td></tr><tr><td>8</td><td>COURSE</td><td></td><td>NEVER</td></tr></table></td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>EH</td><td>I HAVE</td><td>TO</td></tr><tr><td>2</td><td>BECAUSE</td><td></td><td>EVER</td></tr><tr><td>3</td><td>SO</td><td></td><td>SOME</td></tr><tr><td>4</td><td>AND</td><td></td><td>EH</td></tr><tr><td>5</td><td>TEACHER</td><td></td><td>THE</td></tr><tr><td>6</td><td>THINK</td><td></td><td>NO</td></tr><tr><td>7</td><td>HAVE</td><td></td><td>BEEN</td></tr><tr><td>8</td><td>DO</td><td></td><td>LEARNED</td></tr><tr><td>9</td><td>THAT</td><td></td><td>MET</td></tr><tr><td>10</td><td>MAYBE</td><td></td><td></td></tr></table></td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>BECAUSE</td><td>I HAVE</td><td>TO</td></tr><tr><td>2</td><td>EH</td><td></td><td>EH</td></tr><tr><td>3</td><td>SO</td><td></td><td>MANY</td></tr><tr><td>4</td><td>IF</td><td></td><td>SOME</td></tr><tr><td>5</td><td>BUT</td><td></td><td>CHANCE</td></tr><tr><td>6</td><td>AND</td><td></td><td>EVER</td></tr><tr><td>7</td><td>THAT</td><td></td><td>NO</td></tr><tr><td>8</td><td></td><td></td><td>HAVE</td></tr><tr><td>9</td><td></td><td></td><td>YES</td></tr><tr><td>10</td><td></td><td></td><td>THINK</td></tr></table></td></tr>
<tr><td>R1 collocates of I've</td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>EH</td><td>I'VE</td><td>NOT</td></tr><tr><td>2</td><td></td><td></td><td>BEEN</td></tr><tr><td>3</td><td></td><td></td><td>WORKED</td></tr><tr><td>4</td><td></td><td></td><td>EH</td></tr><tr><td>5</td><td></td><td></td><td>NO</td></tr></table></td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>EH</td><td>I'VE</td><td>BEEN</td></tr><tr><td>2</td><td>TIME</td><td></td><td></td></tr></table></td><td colspan="3">Not applicable</td></tr>
</table>

Table 7.11 Use of *I have* and *I've* in LIVT, TESOL and NTESOL

The situation with regard to the use of *I've* is similar to that for the use of *I have*. The pattern *eh + I've + been* was shared by the LIVT and TESOL groups. There are only two instances of *I've* in the NTESOL data, so it is not possible for the WordSmith Tool to generate its patterns of use. However, when we looked at these two instances in the NTESOL group, we found that they were also in the pattern *pronoun + 've + been*, as shown in Table 7.12.

N	Concordance
1	aybe because I learn from I **I've been** learning English from
2	h &eh and children &=cough **they've been** &eh maybe there are

Table 7.12 Instances of *'ve* in NTESOL

Examination of the patterns of *I have* and *I've* in the three groups shows that the participants in these three groups did not have a noticeable difference in their way of using *I have* and *I've* in terms of collocations. In fact, when examining the use of *I have* and *I've* in the reference corpora, it is found that the pattern *connectives + I have + to* and the collocation *I've been* are also very commonly used by the speakers in BASEah and COCAsp. (See Table 7.13.) The reference corpora, of course, show more various patterns than the Taiwanese data, due to their higher occurrences of the use *I have* and *I've*. But still, the findings indicate that in the use of *I have*, the collocation *I have to* is the most common use in all the groups regardless of their different language backgrounds. So is the use of *I've been*. All of the Taiwanese groups and the reference corpora had the collocation *I've been* as their primary use in the use of *I've*.

<table>
<tr><td></td><td colspan="3">BASEah</td><td colspan="3">COCAsp</td></tr>
<tr><td></td><td>Freq.</td><td>%</td><td>Per 1000 words</td><td>Freq.</td><td>%</td><td>Per 1000 words</td></tr>
<tr><td>HAVE</td><td>2122</td><td>100</td><td>5.02</td><td>117818</td><td>100</td><td>6.38</td></tr>
<tr><td>I HAVE</td><td>98</td><td>4.62</td><td>0.23</td><td>9938</td><td>8.44</td><td>0.54</td></tr>
<tr><td>I'VE</td><td>303</td><td>--</td><td><u>0.72</u></td><td>8316</td><td>--</td><td><u>0.45</u></td></tr>
<tr><td>R1 collocates of I have</td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>AND</td><td>I HAVE</td><td>TO</td></tr><tr><td>2</td><td>THAT</td><td></td><td>THE</td></tr><tr><td>3</td><td>THINK</td><td></td><td>NO</td></tr><tr><td>4</td><td>WHAT</td><td></td><td>READ</td></tr><tr><td>5</td><td>BUT</td><td></td><td>TIME</td></tr><tr><td>6</td><td>SO</td><td></td><td>SOME</td></tr><tr><td>7</td><td>KNOW</td><td></td><td>ONLY</td></tr><tr><td>8</td><td>WHICH</td><td></td><td>PUT</td></tr><tr><td>9</td><td>WELL</td><td></td><td>SEEN</td></tr><tr><td>10</td><td>HAVE</td><td></td><td></td></tr></table></td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>AND</td><td>I HAVE</td><td>TO</td></tr><tr><td>2</td><td>THAT</td><td></td><td>N'T</td></tr><tr><td>3</td><td>BUT</td><td></td><td>BEEN</td></tr><tr><td>4</td><td>KNOW</td><td></td><td>NO</td></tr><tr><td>5</td><td>BECAUSE</td><td></td><td>NEVER</td></tr><tr><td>6</td><td>SO</td><td></td><td>THE</td></tr><tr><td>7</td><td>WELL</td><td></td><td>NOT</td></tr><tr><td>8</td><td>WHAT</td><td></td><td>SEEN</td></tr><tr><td>9</td><td>DO</td><td></td><td>AN</td></tr><tr><td>10</td><td>IF</td><td></td><td>SOME</td></tr></table></td></tr>
<tr><td>R1 collocates of I've</td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>AS</td><td>I'VE</td><td>GOT</td></tr><tr><td>2</td><td>WHAT</td><td></td><td>BEEN</td></tr><tr><td>3</td><td>AND</td><td></td><td>SAID</td></tr><tr><td>4</td><td>THAT</td><td></td><td>JUST</td></tr><tr><td>5</td><td>BUT</td><td></td><td>PUT</td></tr><tr><td>6</td><td>WHICH</td><td></td><td>ALREADY</td></tr><tr><td>7</td><td>NOW</td><td></td><td>GIVEN</td></tr><tr><td>8</td><td>WELL</td><td></td><td>MENTIONED</td></tr><tr><td>9</td><td>SORRY</td><td></td><td>DONE</td></tr><tr><td>10</td><td>SEE</td><td></td><td>WRITTEN</td></tr></table></td><td colspan="3"><table><tr><td>N</td><td>L1</td><td>Centre</td><td>R1</td></tr><tr><td>1</td><td>AND</td><td>I 'VE</td><td>BEEN</td></tr><tr><td>2</td><td>THAT</td><td></td><td>GOT</td></tr><tr><td>3</td><td>WHAT</td><td></td><td>NEVER</td></tr><tr><td>4</td><td>KNOW</td><td></td><td>SEEN</td></tr><tr><td>5</td><td>BUT</td><td></td><td>HAD</td></tr><tr><td>6</td><td>BECAUSE</td><td></td><td>EVER</td></tr><tr><td>7</td><td>WELL</td><td></td><td>HEARD</td></tr><tr><td>8</td><td>THINK</td><td></td><td>ALWAYS</td></tr><tr><td>9</td><td>SO</td><td></td><td>DONE</td></tr><tr><td>10</td><td>MEAN</td><td></td><td>SAID</td></tr></table></td></tr>
</table>

Table 7.13 Use of *I have* and *I've* in BASEah and COCAsp

There are not many differences found in terms of the collocations of *I have* and *I've* across groups. However, when examining all of the instances of *I have* and *I've* in the Taiwanese groups, some noticeable differences are found in the use of the word *have*. There are three common use of the word *have* found in the three groups. (See Table 7.14.) One is the use of the word *have* as a general verb for expressing 'possession' as in instance [a] below. The second is the use of *have* as an auxiliary verb with a past participle for expressing recent past or completed action or period of time, as in the instance [b]. The third is the use of *have* with *to*. The collocation *have to* functions like a modal verb for expressing obligation or requirement, as in the instance [c].

[a] &eh how to say that I have **I have** a dream that to make educa
[b] aybe because I learn from I **I've (have) been** learning English
[c] hem can't speak Chinese so **I have to** use English to communic

Table 7.14 Instances of *I have*/*I've* in NTE

When examining how the word *have* was used based on the above classification, it is found that the three groups of participants had their own special tendencies as regards their use of the word *have*. The LIVT participants, when they used the collocation *I have*, tended to use *have* for expressing possession more often than the other two types of use. There are 26 occurrences of possession *have* in the LIVT data, and the use accounts for 41.27% of the total use of *I have* in the LIVT group. The use of *have* for expressing 'past/completed action or time' and 'obligation/requirement' are less frequent in the LIVT data. This finding indicates that when the LIVT participants used the verb *have* in relation to themselves, they were more likely to talk about possession. The TESOL participants had their own preference as well. There are 32 occurrences of the word *have* for expressing 'possession' in the TESOL data, and the use accounts for 37.21% of their total use of *I have*. There are 30 occurrences of *have* for expressing 'obligation/requirement' in the TESOL data, and the use accounts for 34.88% of their total use of *I have* in the TESOL data. The finding indicates that when the TESOL participants used the verb *have* relating to themselves, they were likely to talk about possession and obligation/requirement. The NTESOL participants differed from both the LIVT and TESOL participants. There are 47 occurrences of the word *have* used to express 'obligation/requirement' in the NTESOL data, and the use accounts for 46.53% of their total use of *I have*. The other two kinds of use were not found as frequent in the NTESOL group. The finding indicates that when the NTESOL participants used the verb *have* to relate to themselves, they were more likely to talk about obligation/requirement. (See Table 7.15.)

	LIVT			TESOL			NTESOL		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
I HAVE	63	100	2.18	86	100	2.90	101	100	4.33
Possession *HAVE*	26	**41.27**	0.90	32	**37.21**	1.08	28	27.72	1.20
Aux. *HAVE*	6	9.52	0.21	16	18.60	0.54	8	7.92	0.34
Obligation *HAVE TO (need/must)*	17	26.98	0.59	30	**34.88**	1.01	47	**46.53**	2.02
Unable to classify[15]	14	22.22	0.48	8	9.30	0.27	18	17.82	0.77
	LIVT-NTE			LIVT-TE			TE-NTE		
	LLR	Sig.		LLR	Sig.		LLR	Sig.	
Obligation HAVE TO (need/must)	21.79	0.000	*** -	3.29	0.070	-	9.02	0.003	** -

Table 7.15 Use of *I have* in LIVT, TESOL and NTESOL

What we have found explains why the NTESOL participants had significantly more use of *I have* in comparison with the LIVT and TESOL participants. The NTESOL participants used the collocation *I have (to)* for expressing obligation/requirement much more often than the LIVT and TESOL participants. The obligation *I have (to)* is used 2.02 times per 1000 words in the NTESOL data. This single use of *I have* in the NTESOL data is close to the total frequencies of the use of *I have* in the LIVT and TESOL data. When checking the frequencies of the use of *I have (to)* via log-likelihood ratios, it is found that the LIVT and TESOL participants used the obligation *I have (to)* significantly less often than the NTESOL participants. This finding suggests that our hypotheses 1 and 2 are correct. There are noticeable differences in the use of *I have* between the LIVT and TESOL participants and the NTESOL participants. The differences are likely the result of different discourse preferences between groups. This finding also suggests that the participants with non-English-relevant studying backgrounds (in contrast to the TESOL participants) and who have not been exposed to abundant and authentic English frequently (in contrast to the LIVT participants) had a tendency to talk more about the obligation in learning English with the use of *I have*. One possible explanation is that the participants in the NTESOL group may have different attitudes to English learning from the participants in the TESOL and LIVT groups. It also could be that the NTESOL participants were primed to use *I have (to)* when expressing obligation more strongly than LIVT and TESOL participants. (We will discuss this possibility further in our investigation of the use of *have to* in 7.2.2.) It would be interesting to study the different attitudes to English learning that these EFL learners with

[15] The instances that cannot be classified are the ones in which the speaker paused, and then either started a new sentence or repeated the words *I have* one more time. In the former situation, it is hard to determine which use it is. In the latter situation, we ignored the first use of *I have*, and took the later *I have* for classification.

different language learning background had, or to study how these participants had been primed to use words for expressing obligation, but sadly it falls outside our area of concern in this study. Nevertheless, on the basis of our investigation, we can now explain the two differences between the LIVT and TESOL participants and the NTESOL participants that we mentioned in the early part of this chapter.

In the discussion about why the LIVT and TESOL participants have significantly more use of *I've* in comparison with the NTESOL participants, we assumed that the LIVT and TESOL participants tended to use *have* as auxiliary verbs more often than the NTESOL participants. The findings prove our assumption to be true. There are 27 occurrences of *I + Aux. HAVE-Verbs* found in the LIVT data (0.94 times per 1000 words), and 26 occurrences of *I + Aux. HAVE-Verbs* found in the TESOL data (0.88 times per 1000 words.) In the NTESOL data, this kind of use only occurs 9 times (0.39 times per 1000 words.) The LIVT and TESOL participants used auxiliary *have* in the collocation *I have* for expressing 'past/completed action/time' twice as often as the NTESOL participants. (See Table 7.16.)

	LIVT			TESOL			NTESOL		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
I + Aux. VH	27	100	**0.94**	26	100	**0.88**	9	100	**0.39**
I + Aux. HAVE	6	22.22	0.21	16	61.54	0.54	8	88.89	0.34
I + Aux. 'VE	19	70.37	0.66	10	38.46	0.34	1	11.11	0.04
	LIVT-NTE			LIVT-TE			TE-NTE		
	LLR	Sig.		LLR	Sig.		LLR	Sig,	
I + Aux. VH	**5.98**	**0.015**	*** +**	0.06	0.811	+	**5.01**	**0.025**	*** +**

Table 7.16 Use of *I + aux. have* in LIVT, TESOL and NTESOL

The finding that the LIVT and TESOL participants used *have* in its auxiliary verb function (*I + Aux. HAVE-verbs*) in the context of *I have ('ve)* more often than the NTESOL participants supports for our hypotheses 1 and 2. We will discuss the use of auxiliary *have* with verbs (past participle) later in the section on R1 collocates (7.2.2).

There is another difference worth noting between the LIVT participants and the TESOL and NTESOL participants, and this finding also supports our hypothesis 1 that the EFL learners who go to the UK for study will have noticeable differences in their English use in comparison with the ones in Taiwan. In Table 7.16, we found that the LIVT participants tended to use *have* in abbreviated form *'ve* more often when it is an auxiliary *have*. The use of *I + Aux. 've* occurs 19 times in the LIVT data. It accounts for about 70% of the total use of *I+ Aux. HAVE-Verbs*. The use of *I + Aux. 've* accounts

for less than 50% of the total use of *I + Aux. HAVE-Verbs* in both the TESOL and NTESOL data. This finding indicates that when the LIVT participants used the word *have* for its auxiliary function with *I*, they tended to abbreviate the word. This strong tendency was not found in the TESOL and NTESOL data. Especially in the NTESOL data, there is only one instance of *I've* found. The fact that the use of *I've* was used much more frequently by the LIVT participants is potentially compatible with what has been discussed in chapter 2. In Table 7.12, we can see that the use of *I've* is frequently found in the BASEah and COCAsp corpora. *I've* occurs 0.72 times per 1000 words in the BASEah corpus, and it occurs 0.45 times per 1000 words in the COCAsp corpus. The use of *I've* features frequently in the authentic spoken English to which the LIVT group had very likely been exposed. Therefore, we can say it is very possible the LIVT group was influenced to abbreviate the auxiliary *have* in their spoken English by the input they received. Next, we will investigate the R1 collocates of *have*.

7.2.2 R1 Collocates of *have*

Table 7.17 presents the top 10 most frequent R1 collocates of *have* in the three groups. We can see that the collocations *have to*, *have a*, *have the*, *have I* and *have some* are commonly found in all the Taiwanese groups' lists.

	LIVT	227	7.86	TESOL	264	8.89	NTESOL	250	10.73
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	TO	51	1.77	TO	98	3.30	TO	95	4.08
2	A	36	1.25	A	26	0.88	A	20	0.86
3	THE	16	0.55	THE	21	0.71	EH	17	0.73
4	EH	11	0.38	SOME	12	0.40	THE	13	0.56
5	SUBTITLE	10	0.35	EH	11	0.37	MANY	11	0.47
6	I	8	0.28	I	7	0.24	SOME	9	0.39
7	THIS	6	0.21	MORE	5	0.17	I	9	0.39
8	MORE	6	0.21	EVER	5	0.17	NO	5	0.21
9	SUBTITLES	5	0.17	NO	4	0.13	HAVE	4	0.17
10	SOME	5	0.17	WE	3	0.10	MUCH	3	0.13

Table 7.17 R1 Collocates of *have* in LIVT, TESOL and NTESOL

When we examining these collocates via log-likelihood ratios, it is found that the LIVT participants used the collocation *have to* significantly less frequently than the TESOL and NTESOL participants. The LIVT participants used the collocation *have subtitle* significantly more frequently than the TESOL and NTESOL participants. The LIVT and TESOL participants used the collocation

have many significantly less often than the NTESOL participants. The LIVT participants used the collocation *have this* significantly more frequently than the TESOL participants. (See Table 7.18.)

LIVT - NTE					LIVT - TE					TE - NTE				
Over					Over					Over				
Words	LLR	Sig.			Words	LLR	Sig.			Word	LLR	Sig.		
HAVE SUBTITLE	6.746	0.009	**	+	HAVE SUBTITLE	8.800	0.003	**	+					
					HAVE THIS	4.103	0.043	*	+					
Less					Less					Less				
Word	LLR	Sig.			Word	LLR	Sig.			Word	LLR	Sig.		
HAVE TO	24.561	**0.000**	***	-	**HAVE TO**	13.802	**0.000**	***	-	HAVE MANY	9.227	0.002	**	-
HAVE MANY	12.030	0.001	***	-										

Table 7.18 Log-Likelihood Ratios: *have* in LIVT, TESOL and NTESOL – R1 Collocates

The use of *have subtitle* and *have many* is closely relevant to the content of our discussions during the interviews. We discussed the Chinese subtitles on TV and we also discussed the number of foreign friends that these participants had in Taiwan. Therefore, the use of *have subtitle* and *have many* will not be investigated in further detail. The use of the collocation *have this* occurrs rarely in the three groups. There are only 6 occurrences in the LIVT data (contributed by 4 LIVT participants), but only one occurrence in the TESOL data and one occurrence in the NTESOL data. Therefore, we also exclude it from further investigation. In the following section, we will investigate the use of the collocation *have to*.

HAVE TO

Earlier in our investigation on the use of *I have*, we found that the NTESOL participants tended to use the word *have* with *I* for expressing 'obligation/requirement' twice as frequently as the LIVT and TESOL participants. Both the LIVT and TESOL participants had lower use of obligation *have (to)* in the use of the collocation *I have*. However, in our investigation on the use of *have to*, we found that it is not merely the NTESOL participants that used the obligation sense of *have (to)* in their speech; the TESOL participants also used the collocation *have to* significantly more often than the LIVT participants. The LIVT participants used the word *have* in its use of expressing obligation significantly less frequently than the TESOL and NTESOL participants in general. There are 95

occurrences of *have to* found in the NTESOL data (4.08 times per 1000 words), 98 occurrences of *have to* in the TESOL data (3.30 times per 1000 words) and merely 51 occurrences of *have to* in the LIVT data (1.77 times per 1000 words.)[16] The LIVT participants used the collocation *have to* about half as often as the TESOL and NTESOL participants did. (See Table 7.19.)

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
HAVE	227	100	7.86	264	100	8.89	250	100	10.73
HAVE TO	51	22.47	**1.77**	98	37.12	**3.30**	95	38.00	**4.08**
HAS TO	0	--	0.00	1	--	0.03	0	--	0.00
HAD TO	3	--	0.10	0	--	0.00	3	--	0.13
Patterns of *have to*									

N	L1	Centre	R1
1	YOU	HAVE TO	LEARN
2	WE		USE
3	THEY		WRITE
4	TO		EH
5			GO
6			FOCUS
7			SPEAK
8			HAVE
9			OBTAIN

N	L1	Centre	R1
1	YOU	HAVE TO	EH
2	WE		LEARN
3	THEY		DO
4	DON'T		KNOW
5			USE
6			READ
7			TALK
8			SPEAK
9			TEACH
10			GO

N	L1	Centre	R1
1	YOU	HAVE TO	EH
2	WE		LEARN
3	THEY		USE
4	DON'T		READ
5	IT'S		DO
6			WRITE
7			GO
8			TALK
9			STUDY
10			CHAT

Table 7.19 Use of *have* and *ha* to* in LIVT, TESOL and NTESOL

Despite the fact that the TESOL and NTESOL participants had a significantly higher frequency of the use of *have to* in their speech, how these participants used the collocation *have to* did not differ from that of the other groups. As seen in Table 7.19, the patterns of the use of *have to* in all the three groups were very much alike. The patterns *you/we + have to + learn/use* were shared by all the groups of participants. The significantly more frequent use of *have to* in the TESOL and NTESOL data, as we discussed in connection with the use of *I have*, may be a result from a tendency in which the TESOL and NTESOL participants tended to express 'obligation/requirement' more often than the LIVT participants. This may stem from a different attitude towards English learning which had influenced their significantly high use of *have to* in the TESOL and NTESOL groups. However, it is also possible that the attitudes to English learning may not differ greatly between the three groups, but that when expressing obligation, the participants in the TESOL and NTESOL groups had been overly primed to use the phrase *have to*, with the consequence that *have to* is the expression that they knew the best for expressing obligation. The participants in the LIVT group might have a similar amount of

[16] We also checked other forms of *have* (*has, had*) with *to*, but the frequencies of such use are few. Therefore, we will not include them in the discussion of the use of *have to*.

use of obligation in their talk, but they were likely to have been primed in more ways for expressing obligation due to their comparatively greater exposure to different varieties of English input. To test whether this assumption is correct, a thorough discourse analysis of these participants' use of obligation would be required. However, this falls outside the focus of the current research. Our focus is on whether the different linguistic contexts (the experience of studying in an English-speaking country/ the experience of receiving English-relevant educational training) results in noticeable different language use among the Taiwanese EFL learners. With this focus in mind, we note that the greater use of the words *have to* by the TESOL and NTESOL participants and the smaller use of the words *have to* by the LIVT participants indicate that our hypothesis 1 for English use is correct. In the next section, we will investigate the colligation (lexico-grammatical features) *HAVE-Verbs* with *past participles*.

HAVE-Verbs + Verbs (Past Participle)

In the investigation of *I have*, we found that the LIVT and TESOL participants tended to use *have* as an auxiliary verb in the collocation *I have* twice as often as the NTESOL participants. When we checked all of the total use of *HAVE-verbs* with *verbs (past participles)*, we found that there are 20 occurrences of *HAVE-verb + V/PP* in the LIVT data (0.69 times per 1000 words), 23 occurrences of *HAVE-verb + V/PP* in the TESOL data (0.77 times per 1000 words) and 13 occurrences of *HAVE-verb + V/PP* in the NTESOL data (0.56 times per 1000 words.) (See Table 7.20.)

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
VH (Have-Verbs)	261	100	9.04	303	100	10.21	264	100	11.33
VH+ Verbs/PP	20	7.66	**0.69**	23	7.59	**0.77**	13	4.92	**0.56**
VH + PP	13	4.98	0.45	21	6.93	0.71	7	2.65	0.30
VH+V[17]	7	2.68	0.24	2	0.66	0.07	6	2.27	0.26
R1 Collocates of VH	BEEN 4 HAVE 3 WORKED 2 ASK 2			BEEN 7 LEARNED 3 MET 2			HAVE 4 BEEN 3 TAUGHT 2		

Table 7.20 Use of *have* and *ha* to* in LIVT, TESOL and NTESOL

[17] The use of *HAVE-Verb + Verb* in these groups is actually the same as the use of *HAVE-Verb + Past participle*. The participants did not pronounce the verb in its past participle form. They pronounced its simple verb form (e.g. the verbs *have* and *ask* in the LIVT list of R1 collocates.) This is possibly the result of different primings at the level of phonology. This may point to insufficient priming for the colligation with participle in the R1 position. Since the two kinds of use (*HAVE-Verb + Verb* and *HAVE-Verb + Past participle*) are for the same purpose (expressing 'past/completed action/time'), we treated these two patterns as one in our discussion.

The use of *HAVE-verb + V/PP* was used more often by the LIVT and TESOL participants in comparison with by the NTESOL participants. The TESOL participants even used this usage more often than the LIVT participants. However, when examining the frequencies of the use of *HAVE-verb + V/PP* in the three groups via log-likelihood ratios, it is found the differences in terms of the frequencies of the use of *HAVE-verb + V/PP* does not reveal significance. (See Table 7.21.) The words used in the patterns are also quite similar to each other in these groups. (See Table 7.20.)

Have- Verb + PP/V	LLR	Sig.		
LIVT - TE	0.13	0.714		-
LIVT - NTE	0.37	0.540		+
TE - NTE	0.92	0.337		+

Table 7.21 Log-Likelihood Ratios: *VH+V/PP* in LIVT, TESOL and NTESOL

In our previous discussion of the use of auxiliary *have* in the collocation *I have*, we found that it is not that the NTESOL participants did not use the *Aux. have* with *verbs (PP)* to express 'past/completed action or time' in their spoken English. It is that the NTESOL participants did not use this kind of expression as frequently as the LIVT and TESOL groups when they were talking about themselves. Here, in our investigation on the total use of *HAVE-verbs + V/PP* to express 'past/completed action/time', the NTESOL participants did not differ significantly from the LIVT and TESOL participants, which tells us that the significant difference in the use of *I have* is very likely to be an issue of different discourse preferences, rather than lexico-grammatical differences. These findings serve as an evidence for supporting our hypotheses that 1) the EFL learners who go to the UK would have noticeable differences in their English use in comparison with the ones in Taiwan, and 2) the EFL learners who major in English-relevant subjects in Taiwan would have noticeable differences in their English use in comparison with the ones with non-English-relevant subjects in Taiwan. However, the reason behind this noticeable difference is not relevant to our concern about the influence caused by the different language input on people's language use, but the reason is more relevant to the social/educational differences of these EFL participants as discussed at the beginning of chapter four.

7.2.3 Summary of the use of *have*

In our investigation on the use of English word *have* (and *HAVE-Verbs* in general), some noticeable differences between the groups were found. The findings are presented in Table 7.22 and Table 7.23.

Findings that Agree with Hypothesis 1			
Noticeable Differences	Collocations	The types of differences	Descriptions OR noteworthy findings/observation
The LIVT participants used the collocations **significantly more frequently** than the NTESOL participants.	*I've*	Lexico-grammatical Difference	The LIVT participants were prone to use *have* in its auxiliary verb function (I + Aux. HAVE-verbs) in the context of *I have('ve)* more often than the NTESOL participants. The LIVT participants were prone to abbreviate the word *have* more often than the TESOL and NTESOL participants when they use the word *have* in its auxiliary function with *I*.
The LIVT participants used the collocations **significantly less frequently** than the NTESOL participants.	*I have/ I have (to)*	Discourse preference	The LIVT participants did not use the obligation *I have (to)* as often as the NTESOL participants did in their use of *I have*.
The LIVT participants used the collocations **significantly less frequently** than the TESOL and NTESOL participants.	*have to*	Discourse preference	The LIVT participants tended not to use the word *have* in its use of expressing obligation as often as the TESOL and NTESOL participants in general.

Table 7.22 Findings in the use of *have* that support Hypothesis 1

With regard to our hypothesis 1, we have found that the LIVT participants used the word *have* in its auxiliary function in the context of use of *I have* more often than the NTESOL participants, and they also abbreviated the auxiliary *have* with the pronoun *I* more often than the participants in the TESOL and NTESOL groups. The latter use is possibly influenced by their English input, because we also observed that abbreviated *have* with *I* is used frequently in the reference corpora. We also found that the LIVT participants used the obligation *have (to)* less frequently than the TESOL and NTESOL participants. The difference of discourse preferences may be caused by the different social/educational

factors which had influenced the participants' viewpoints towards English learning. It is also possible that the input factor had influenced the TESOL and NTESOL participants so that they were primed more strongly to use *have (to)* for expressing obligation, or that the LIVT participants were primed weakly to use *have (to)* for this purpose. Either way, these findings suggest that our hypothesis 1 is correct.

Findings that Agree with Hypothesis 2			
Noticeable Differences	Collocations	The types of differences	Descriptions OR noteworthy findings/observation
The TESOL participants used the collocations **significantly more frequently** than the NTESOL participants.	*I've*	Lexico-grammatical Difference	The TESOL participants were prone to use *have* in its auxiliary verb function *(I + Aux. HAVE-verbs)* in the context of *I have('ve)* more often than the NTESOL participants.
The TESOL participants used the collocations **significantly less frequently** than the NTESOL participants.	*I have/ I have (to)*	Discourse preference	The TESOL participants did not use the obligation *have (to)* as often as the NTESOL participants did in context of their use of *I have*.

Table 7.23 Findings in the use of *have* that support Hypothesis 2

With regard to our hypothesis 2, we found that the TESOL participants, like the LIVT participants, used the auxiliary *have* in the context of *I have* more often than the NTESOL participants. They however used the obligation *have (to)* with the pronoun *I* less frequently than the NTESOL participants. The findings of these noticeable differences support our second hypothesis. In the next section, we will investigate the use of Mandarin Chinese word 有 (yǒu).

7.3 The Use of 有(yǒu) [to have/exist]

In this section, we will investigate whether there are any noticeable differences in the use of the word 有 (yǒu) in the three groups' spoken Mandarin Chinese. In our third and fourth hypotheses, we hypothesised there would be no noticeable differences between these Taiwanese groups due to their shared Mandarin Chinese input in Taiwan. However, from our previous investigation on the use of 是 (shì) [BE], we found that our hypotheses 3 and 4 are incorrect. There were noticeable differences between the groups in their Mandarin use. Furthermore, several findings, such as the TESOL participants' use of *is that* and the linking 是 (shì) [BE] in context of use of 是你 [BE you] suggest there is a possibility that the participants' English primings of language use has influenced their Mandarin primings. In this section, we will continue testing our hypotheses 3 and 4 on the use of 有 (yǒu) [to have/exist].

7.3.1 L1 Collocates of 有 (yǒu) [to have/exist]

Table 7.24 presents the top 10 most frequent L1 collocates of 有 (yǒu). The negative marker 没[no/not], the verb 是[BE], the modal verb 会[can], the pronoun 我[I], and the adverbs 还[still] and 就是[just] are L1 collocates of 有(yǒu) [to have/exist] shared in all the three groups' lists. There are however several collocates that appear only in one or two groups' top 10 most frequent L1 collocate lists. For instance, the words, the verb 比较 [compare], the adverb 还是[still/nevertheless] and the pronoun 你[you], only appear in the LIVT group's list.

In Table 7.25, we can see that the LIVT participants used the collocation 还是 有 [still/nevertheless have/exist] significantly more frequently than the TESOL and NTESOL participants, and that the LIVT participants used the collocation 还 有 [still/in addition have/exist] significantly less frequently than the TESOL and NTESOL participants. The TESOL participants used the collocations 也 有 [also have/exist], 很 有 [very have/exist] and 对 有 [yes/to have/exist] significantly more often than the LIVT and NTESOL participants. The TESOL participants used the collocation 没 有 [no/not have/exist] significantly less often than the LIVT and NTESOL participants. There are noticeable differences in the frequencies of the use of collocations 嗯 有[eh have/exist] and 有 有[repetition of have/exist] between groups. However, as mentioned previously,

we will not be investigating the discourse fillers 对 有 [yes/to have/exist] and 嗯 有[eh have/exist] , the repetition 有 有 and the negation 没 有 [no/not have/exist] in this study.

	LIVT	870	22.56	TESOL	870	24.64	NTESOL	1059	25.14
N	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words	L1	Freq.	per 1000 words
1	没	273	7.08	没	207	5.86	没	337	8.00
2	是	60	1.56	还	52	1.47	会	81	1.92
3	会	59	1.53	是	48	1.36	还	52	1.23
4	还	29	0.75	会	47	1.33	是	52	1.23
5	有	22	0.57	很	28	0.79	就是	29	0.69
6	比较	20	0.52	也	24	0.68	我	20	0.47
7	就是	19	0.49	对	20	0.57	它	20	0.47
8	你	18	0.47	就是	19	0.54	都	16	0.38
9	我	17	0.44	都	18	0.51	很	16	0.38
10	还是	16	0.41	我	18	0.51	嗯	16	0.38

Table 7.24 L1 Collocates of 有(yǒu)[to have/exist] in LIVT, TESOL and NTESOL

LIVT - NTE						LIVT – TE						TE - NTE					
Over						Over						Over					
Word	With	LLR	Sig.			Word	With	LLR	Sig.			Word	With	LLR	Sig.		
还是	**有**	**10.947**	**0.001**	***	+	没	有	4.216	0.040	*	+	很	有	5.771	0.016	*	+
你	有	6.027	0.014	*	+	**还是**	有	**3.877**	**0.049**	*	+	也	有	5.546	0.019	*	+
有	有	3.932	0.047	*	+							对	有	4.478	0.034	*	+
Less						Less						Less					
Word	With	LLR	Sig.			Word	With	LLR	Sig.			Word	With	LLR	Sig.		
嗯	有	14.672	0.000	***	-	也	有	17.668	0.000	***	-	没	有	12.674	0.000	***	-
还	**有**	**4.346**	**0.037**	*	-	有	有	10.657	0.001	**	-						
						还	**有**	**8.808**	**0.003**	**	-						
						很	有	8.072	0.004	**	-						
						对	有	4.338	0.037	*	-						

Table 7.25 Log-Likelihood Ratios: 有(yǒu)[have/exist] in LIVT, TESOL and NTESOL – L1 collocates

The finding that the TESOL participants used the adverbs 也[also] and 很 [very] with 有 (yǒu) [to have/exist] significantly more often than the LIVT and NTESOL participants is interesting.

However, our main purpose of this study is to find out whether different English learning backgrounds will cause noticeable differences in language use. In our previous discussions of Mandarin use in chapter 6, we found that the English primings may influence the participants' Mandarin primings. Therefore, if possible, we can test whether our tentative finding is supported by investigation of the use by the three groups of 有 (yǒu) [to have/exist]. Based on this reason, we will select the collocations which are more likely to meet our purpose. There are two collocations 还是 [still/nevertheless have/exist] and 还有 [still/in addition have/exist] that seem to be quite similar in their composition of words (还[still], 是[BE] and 有[have/exist]), and the finding that the LIVT participants used the former one significantly more often and the latter one significantly less often than the TESOL and NTESOL participants catches our attention. Therefore, we will investigate the use of both the collocations 还有 [still/in addition have/exist] and 还是 有 [still/nevertheless have/exist] in detail.

还有(hái yǒu) [still/in addition have/exist]

In the previous section, we noted that the LIVT participants used the collocation 还有[still/in addition have/exist] significantly less frequently than the TESOL and NTESOL participants. (See Table 7.26.)

还(hái) 有(yǒu)				
	LLR	Sig.		
LIVT - TE	8.808	**0.003**	**	-
LIVT - NTE	4.346	**0.037**	*	-
TE - NTE	0.985	0.321		+

Table 7.26 Log-Likelihood Ratios: 还有(háiyǒu) [still have/exist] in LIVT, TESOL and NTESOL

There are 52 occurrences of the use of 还有[still/in addition have/exist] in the TESOL data (1.47 times per 1000 words) and also 52 occurrences of the use of 还有[still/in addition have/exist] in the

NTESOL data (1.23 times per 1000 words.) There are 29 occurrences of the use of 还有[still/in addition have/exist] in the LIVT data (0.75 times per 1000 words.) (See Table 7.27.)

	LIVT		**TESOL**		**NTESOL**	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
有(yǒu)	870	22.56	870	24.64	1059	25.14
还有(hái yǒu)	29	**0.75**	52	**1.47**	52	**1.23**

Table 7.27 Frequencies of the use of 还有(hái yǒu) [still/in addition have/exist] in LIVT, TESOL and NTESOL

The LIVT participants used the collocation 还有[still/in addition have/exist] nearly twice as infrequently as the TESOL participants. The LIVT participants also used the collocation 还有[still/in addition have/exist] slightly less often than the NTESOL participants. When investigating on all of the instances of how these participants used 还有 [still/in addition have/exist], two main kinds of use of the combination 还[still] and 有[to have/exist] are found.

The first kind of use is to use the word 有(yǒu)[to have/exist] as a verb, and the word 还[still] is used as an adverb for modifying the verb 有(yǒu)[to have/exist]. (See Example LIVT 7-1a in Table 7.28.) If the adverb 还[still] is removed, the sentence is still comprehensible. (See Example LIVT 7-1b.)

LIVT 7-1a	我 们 都 不 会 想 到 , 其实 **还 有** 其他 的 方 面 可以 写 这 个 东西。
	Wǒmen dōu bú huì xiǎngdào , qíshí hái yǒu qítā de fāngmiàn kéyǐ xiě zhè gè dōngxi。
	All of us won't realise that actually **there are still** other aspects [for us] to write about this thing.
LIVT 7-1b	我 们 都 不 会 想 到 , 其实 **有** 其他 的 方 面 可以 写 这 个 东西。
	Wǒmen dōu bú huì xiǎngdào , qíshí yǒu qítā de fāngmiàn kéyǐ xiě zhè gè dōngxi。
	All of us won't realise that actually **there are** other aspects [for us] to write about this thing.
NTE7-1a	[我读] 推理 小 说 , 然 後 奇 幻 小 说 也 有 , 然後 **还 有** 村 上 春 树 的 书 ,
	[wǒdú] tuīlǐ xiǎoshuō , ránhòu qíhuàn xiǎoshuō yě yǒu , ránhòu hái yǒu cūnshàngchūnshù de shū ,
	I read detective novels, and then also fantasy fictions, and then**, and also** Murakami Haruki's books.
NTE7-1a	***[我读] 推理 小 说 , 然 後 奇 幻 小 说 也 有 , 然後 **有** 村 上 春 树 的 书 ,
	***[wǒdú] tuīlǐ xiǎoshuō , ránhòu qíhuàn xiǎoshuō yě yǒu , ránhòu yǒu cūnshàngchūnshù de shū ,
	*I read detective novels, and then also fantasy fictions, and then**, have** Murakami Haruki's books.

Table 7.28 Examples of the use of 还有(hái yǒu) [still/in addition have/exist]

The second kind of use is to use the word 有 (yǒu) [have/exist] as a suffix to an adverb 还 (hái) [still] and this combination can function as a connective, which means 'and also' if the combination 还有 (hái yǒu) is used between noun phrases/verb phrases. (See Example NTE 7-1a in Table 7.28.) This combination means 'in addition' when it is used at the beginning of the sentences. In this kind of use, the word 还 (hái) [still] cannot be omitted. If the word 还 (hái) [still] is removed, the sentence will be incomprehensible. (See Example NTE7-1b in Table 7.28.)

	LIVT		TE		NTE	
	Freq.	Per 1000 words	Freq.	Per 1000 words	Freq.	Per 1000 words
还有(hái yǒu)	29	0.75	52	1.47	52	1.23
有(yǒu) as a verb (still have)	4	0.10	6	0.17	5	0.12
有(yǒu) as a suffix (and; in addition)	25	**0.65**	46	**1.30**	47	**1.12**

	LIVT-NTE			LIVT-TE			TE-NTE		
	LLR	Sig.		LLR	Sig.		LLR	Sig.	
有(yǒu) as a suffix (and; in addition)	**5.03**	**0.025**	* -	**5.73**	**0.017**	* -	0.56	0.456	+

Table 7.29 The use of the word 有(yǒu) in 还有(hái yǒu) in LIVT, TESOL and NTESOL

We found that the LIVT participants used the second kind of use significantly less often than the TESOL and NTESOL participants. There are 25 instances of the use of the suffix 有 (yǒu) [have/exist] with the word 还 (hái) [still] in the LIVT data (0.65 occurrences per 1000 words), 46 instances of the use of the suffix 有 (yǒu) [have/exist] with the word 还 (hái) [still] in the TESOL data (1.30 occurrences per 1000 words) and 47 instances of this usage in the NTESOL data (1.12 occurrences per 1000 words.) (See Table 7.29.)

The finding suggests that the participants in the LIVT group did not use the collocation 还有 (hái yǒu) in its connective function as often as the participants in the TESOL and NTESOL groups. The choice to use 还有 (hái yǒu) as a connective more often or less often may a difference of discourse preference. The LIVT participants might not use the connective 还有 (hái yǒu) for talking about parallel things or ideas as often as the TESOL and NTSOL participants. However, this explanation, although it is not impossible,

would be very strange if we considered that the LIVT and the TESOL and NTESOL participants all answered the same set of questions, and they had all received similar Mandarin input since their childhood. It is difficult to explain why the LIVT participants would either talk less about the parallel things/ideas or avoid using the connective 还有 (háiyǒu) for this purpose, since what we observed from the figures of this use in the TESOL and NTESOL groups suggests the two kinds of use were commonly shared by the majority of our Taiwanese participants (mostly the participants in the TESOL and NTESOL groups.) It is not easy to imagine a factor that would make the LIVT participants tended not to talk about parallel things/ideas or use the connective 还有 (háiyǒu) for talking about parallel things/ideas. The most distinct difference between the participants in the LIVT group and in the TESOL and NTESOL groups is the different degree of their English exposure. That is to say, it is very likely that the LIVT participants were influenced by the way their English input had primed them in their use of English and that their primings of English use, to some extent, had also influenced their primings of Mandarin use. There are several possibilities. One of the possibilities is that, the Mandarin context where the connective 还有 (háiyǒu) is used may be a context where the equivalent English connectives are rarely used. If this is the case, the LIVT participants might be primed to avoid using the connective (还有 (háiyǒu)) in such a context. Another possibility is that their English primings with regard to the connectives might have caused the LIVT participants to use other kinds of Mandarin connectives more frequently instead of the connective 还有 (háiyǒu) to organise their texts/ideas. Of course these possibilities need a thoroughout discourse investigation into how these participants used connectives in both languages, so that there is more evidence to test whether these assumptions are correct or not. Unfortunately, this is not the focus of the current study, but it would be an important area to explore in the future. Nevertheless, what we have found in the use of 还有 (háiyǒu) [still/in addition have/exist] is further evidence that null hypothesis 3 is incorrect.

还是有(háishì yǒu) [still/nevertheless have/exist]

With regard to the use of 还 是 有 [still BE have], it is found that the LIVT participants used the collocation 还 是 有 [still BE have] significantly more frequently than the TESOL and NTESOL participants. (See Table 7.30.)

还 是 有(hái shì yǒu)				
	LLR	Sig.		
LIVT - TE	3.877	**0.049**	*	+
LIVT - NTE	10.947	**0.001**	***	+
TE - NTE	1.618	0.203		+

Table 7.30 Log-Likelihood Ratios: 还是有 (hái shì yǒu) in LIVT, TESOL and NTESOL

There are 16 occurrences of the use of 还是 有 [still BE have] in the LIVT data (0.41 times per 1000 words) and 6 occurrences of the use of 还是 有 [still BE have] in the TESOL data (0.17 times per 1000 words.) There are merely 3 occurrences of the use of 还是 有 [still BE have] in the NTESOL data (0.07 times per 1000 words.) (See Table 7.31.) Here we can see that the differences in frequencies between groups are quite distinct. The LIVT participants used the collocation 还是 有 [still BE have] over twice as often as the TESOL participants. The LIVT participants used the collocation 还是 有 [still BE have] almost 6 times as often as the NTESOL participants. Even the TESOL participants showed double the use of the collocation 还是 有 [still BE have] in comparison with the NTESOL participants.

	LIVT			**TESOL**			**NTESOL**		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
有(yǒu)	870	100	22.56	870	100	24.64	1059	100	25.14
还 是 有(hái shì yǒu)	16	1.84	**0.41**	6	0.69	**0.17**	3	0.28	**0.07**

Table 7.31 Use of 还是有 (hái shì yǒu) in LIVT, TESOL and NTESOL

The collocation 还是 有 [still BE have] is a combination of the adverb 还是 [still] and the verb 有 (yǒu) [have/exist]. The adverb 还是 [still] is used as a separate unit to modify the verb 有 (yǒu) [have/exist]. Therefore, the meaning of this combination is close to 'it still has/exists…/there is still …' The adverb 还是 [still] itself is a combination of the adverb 还 [still] with the word 是 [BE]. As we noted in chapter 6, the word 是 [BE] is often used as a suffix to an adverb or a connective, and the word also adds emphasis to an expression. Therefore, the use of 还是 有 [still BE have] 有, as we see in Examples LIVT7-2a and TE7-1a in Table 7.32, has a sense of 'nevertheless/no matter how…' in English. Similar to 还 有 [still have/exist] (the first kind of use, in which the word 有 is a verb and the word 还 is an adverb), the adverb 还是 [still] can be removed without interfering with the listener's understanding of the sentence. However, the sense of emphasis is also gone. (See Example LIVT7-2b and TE7-1b.)

LIVT 7-2a	因 为 它 [繁 体字] 毕竟 **还 是 有** 它 的 美 。
	Yīnwèi tā [fántǐzì] bìjìng **háishì yǒu** tā de měi。
	Because it [the traditional Chinese characters] **still has** its own beauty after all. [no matter what you think of it.]
LIVT 7-2b	因 为 它 [繁 体字] 毕竟 **有** 它 的 美 。
	Yīnwèi tā [fántǐzì] bìjìng **yǒu** tā de měi。
	Because it [the traditional Chinese characters] **has** its own beauty after all.
TE7-1a	我 觉 得 它 [写 作] 还 是, 也 还 是 , 也 **还 是 有** 重 要 性。
	Wǒ juédé tā [xiězuò] háishì , yě háishì , yě **háishì yǒu** zhòngyàoxìng
	I think it [writing] **still has** its importance. [no matter what you think of it.]
TE7-1b	我 觉 得 它 [写 作], 也 , 也 **有** 重 要 性。
	Wǒ juédé tā [xiězuò] , yě yě **yǒu** zhòngyàoxìng
	I think it [writing] also **has** its importance.

Table 7.32 Examples of the use of 还是有(hái shì yǒu) [still BE have]

We examined all of the instances and found that despite the significant frequent occurrences of 还是 有 in the LIVT data, the way the collocation 还是 有 [still BE have] was used does not differ very much between the groups in respect of the collocates along with this expression. There were no special patterns used more often by any group of participants in particular. The finding suggests that the LIVT participants tended to use this collocation 还是 有 [still BE have] more often than the TESOL and NTESOL participants for providing an alternative idea to another contrary idea. This difference seems to be a result of the different discourse preferences. However, as we commented in

the previous section on the investigation into the use of 还 有 [still/in addition have/exist], why this particular tendency occurs in the LIVT group and not in the TESOL and NTOESL groups is a question worth studying. Especially if we consider our investigations into the use of 是 说 [BE say] in 6.3.2 and the use of 我 就[I just] in 5.2.2, we found that the LIVT participants had a consistent tendency to use certain Mandarin words (就 [just] in the context of 我 就 [I just] and 是 [BE] in the context of 是 说[BE say]) for certain discourse purposes (e.g. adding emphasis/modality) significantly more often than the TESOL and NTESOL participants. Whether or not the different English input/learning experience/training influenced participants' discourse choices, the significantly frequent use of 还是 有 [still BE have] in the LIVT data shows that null hypothesis 3 is incorrect.

Next, we will investigate the use of R1 collocates of the word 有(yǒu)[have/exist].

7.3.2 R1 Collocates of 有(yǒu)[have/exist]

Table 7.33 presents the top 10 most frequent R1 collocates of 有 (yǒu). We can see that the noun/verb 帮助 [help], the number 一 [one], the classifiers 些 / 一些 [some] and the adverb/pronoun 什么 [what/something] are R1 collocates of 有 (yǒu) shared by all the three groups. There are several collocates that appear only in one or two groups' top 10 most frequent L1 collocate lists. For instance, the noun 兴趣 [interest/hobby], the determiner 那 [that] and the adverb 就是 [just] appear in the LIVT and TESOL groups' top 10 most frequent R1 collocate lists. The adverb 很 [very] appears only in the NTESOL group's top 10 list. We therefore now examine these collocates via log-likelihood ratios.

	LIVT	870	22.56	TESOL	870	24.64	NTESOL	1059	25.14
N	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words	R1	Freq.	per 1000 words
1	帮助	66	1.71	帮助	65	1.84	帮助	95	2.26
2	一	42	1.09	些	43	1.22	一些	47	1.12
3	什么	41	1.06	什么	40	1.13	什么	46	1.09
4	些	33	0.86	一些	30	0.85	些	40	0.95
5	兴趣	28	0.73	那	24	0.68	一	40	0.95
6	一些	27	0.70	一	23	0.65	很	30	0.71
7	那	23	0.60	办法	21	0.59	办法	28	0.66
8	有	23	0.60	兴趣	18	0.51	没	28	0.66
9	就是	22	0.57	很多	17	0.48	它	23	0.55
10	没	21	0.54	就是	15	0.42	在	18	0.43

Table 7.33 R1 Collocates of 有 (yǒu) in LIVT, TESOL and NTESOL

LIVT - NTE						LIVT - TE						TE - NTE					
Over						Over						Over					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
有	兴趣	10.537	0.001	**	+	有	一	4.086	0.043	*	+	有	兴趣	3.945	0.047	*	+
Less						Less						Less					
With	Word	LLR	Sig.			With	Word	LLR	Sig.			With	Word	LLR	Sig.		
有	一些	3.855	0.050	*	-	有	很多	6.497	0.011	*	-	有	它	5.168	0.023	*	-
有	很	3.863	0.049	*	-												
有	它	4.182	0.041	*	-												

Table 7.34 Log-Likelihood Ratios: 有 (yǒu) in LIVT, TESOL and NTESOL – R1 collocates

In Table 7.34, we noted that the LIVT and TESOL participants used the collocation 有 兴趣 [have interest] significantly more often than the NTESOL participants, and they used the collocation 有 它 [have it] significantly less often than the NTESOL participants. We also noted that there are several collocations that were used by one particular group of participants more than by the other groups of participants. For instance, the NTESOL participants used the collocation 有 一些 [have some] significantly more often than the LIVT participants, and the TESOL participants used the collocation 有 很多 [have many] significantly more often than the LIVT participants. The collocation 有 兴趣 is the combination of the verb 有 [have] and the noun 兴趣 [interest]. The collocation 有 一些 is the combination of the verb 有 [have] and 一些 [some: number + quantifier], which is usually accompanied by a noun. The collocation 有 很多 is a combination of the verb 有 [have] and 很多 [many], which is also often accompanied by a noun. The use of 有 它 is accompanied by 的 [DE/of] frequently in the groups. The collocation 有 它的 means 'have its + noun phrase.' The frequent use of the collocations in the pattern *有 + noun/noun phrase* therefore leads us to investigate the R1 colligates (grammatical features) of 有 (yǒu) [have].

In the investigation on all the instances of the word 有 (yǒu) in terms of its R1 colligates, one interesting finding is found worth noting. To explain this, there is a need to describe more specifically how the word 有 (yǒu) is used in Mandarin Chinese.

In Mandarin Chinese, as already noted, the word 有 (yǒu) means 'have/exist'. In Ross and Ma's book, *Modern Mandarin Chinese Grammar (2ed.)* (2014), they categorise expressions involving the word 有(yǒu) in terms of four basic purposes. First, it is used for expressing duration. This is usually in the pattern *verb* + *object* + 有 (yǒu) + *duration*, as shown in Example (A). The word 了, which is a grammatical particle meaning 'past/completion', is sometimes used in this expression. The word 了, is usually used in the end of the sentence.

(A)	我 等 他 **有** 三 小 时 了。
	Wǒ děng tā **yǒu** sān xiǎoshí le。
	I have been waiting for him for three hours.

The second purpose is for expressing possession. To say that someone/something possesses something, the word 有(yǒu) is used after the possessor. It is usually followed by a noun phrase. (See Example (B).)

(B)	他 有 一台 车 。
	Tā yǒu yìtái chē。
	He has a car.

The third purpose is for expressing existence, as shown in Example (C). The word 有(yǒu) is also followed by a noun phrase.

(C)	这 附近 有 一座 停 车 场 。
	Zhè fùjìn yǒu yízuò tíngchēchǎng。
	There is a parking lot nearby.

The distinction between possession and existence can be quite blurred. The use of 有(yǒu) in a sentence often can be interpreted either as possession or existence. It is determined by whether the subject is perceived to be a possessor or a location. (Ross and Man, 2014: 64) (See Sample (D).)

(D)	这 个 图书 馆 **有(yǒu)** 很 多 书 。
	Zhè ge túshūguǎn **yǒu** hěnduō shū。
	This library has many books. (Possession)
	There are many books in this library. (Existence)

The fourth purpose is to express temperature. This is an extension of the expression of existence. If it is used for expressing temperature, the use of 有 (yǒu) is interchangeable with the word 是 [BE].

(D)	今 天 的 温 度 **有** 三 十 度。= 今 天 的 温 度 **是** 三 十 度。
	Jīntiān de wēndù **yǒu** sānshí dù。= Jīntiān de wēndù **shì** sānshí dù。
	Today's temperature is 30

The word 有 (yǒu) can also be found in yes-no questions indicating the completion of an action and talking about the past. There are two kinds of expression for this purpose. The first kind involves the use of the word 了, and the word 有 (yǒu) is accompanied with the negative marker 没, as shown in Example (E). It is usually answered with the action verb followed by the word 了.

(E)	Q: 你 买 **了** 票 **没 有** ?	A: 买 **了**。
	Q: Nǐ mǎi **le** piào **méi yǒu**?	A: mǎi **le**。
	Q: Have you bought the ticket?	A: [I] bought [it].

The second kind of expression involves a fixed term 有没有 (yǒu méi yǒu) (means 'have, (or) have not') and omits the use of 了[1e], as shown in Example (F). It is usually answered with the word 有 (yǒu) or its negative form 没有 (méi yǒu), instead of the action verb itself. This kind of use, as pointed out by Ross and Ma (2014: 246), is used in Taiwan particularly.

(F)	Q:你 **有 没 有** 买 票 ?	A: **有**。 / **没 有**。
	Q: Nǐ **yǒu méi yǒu** mǎi piào ?	A: **yǒu**。/ **méi yǒu**。
	Q: Have you bought the ticket?	A: [Yes, I] have. / [No, I] have not.

Originally the second kind of use in expressing completed action/past was for interrogative use. Nowadays, though, it is used in declarative sentences as well. We will see this usage later in the three Taiwanese groups.

To sum up, 有 (yǒu) is often used for expressing (a) duration (accompanied by duration phrases), (b) possession/existence (accompanied by noun phrases) and (c) completed action (accompanied by verb phrases.) Therefore, we need to check how the Taiwanese participants used the nouns and verbs after the word 有(yǒu).

有 (yǒu) + Nouns

When examining the use of 有 (yǒu) with R1 colligate nouns, it is found that the participants were not very different from each other in terms of the frequencies of the use of R1 nouns. There are 211 occurrences of 有 (yǒu) + *NOUN* in the LIVT data (5.47 times per 1000 words), 206 occurrences of 有 (yǒu) + *NOUN* in the TESOL data (5.83 times per 1000 words), and 275 occurrences of 有(yǒu) + *NOUN* in the NTESOL data (6.53 times per 1000 words.) (See Table 7.35.) The LIVT participants used the pattern 有(yǒu) + *NOUN* slightly less frequently than the TESOL and NTESOL participants, but this difference showed no significance in log-likelihood ratios. (See Table 7.36.)

	LIVT			TESOL			NTESOL		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
有(yǒu)	870	100	22.56	870	100	24.64	1059	100	25.14
有(yǒu) + Nouns	211	24.25	5.47	206	23.68	5.83	275	25.97	6.53
Nouns with 有(yǒu)	bāng zhù **帮助**	**66**		bāng zhù **帮助**	**65**		bāng zhù **帮助**	**95**	
	xìng qù **兴趣**	**28**		rén 人	12		bàn fǎ **办法**	**28**	
	bàn fǎ **办法**	**18**		yì yì 意义	9		rén 人	15	
	rén 人	12		gǎn jué 感觉	6		xìng qù **兴趣**	**10**	
	shí hou 时候	6		jī huì 机会	6		zì mù 字幕	6	
	nèi róng 内容	4		xū yào 需要	4				
	gǎn jué 感觉	4		dào lǐ 道理	4				
				zuò wén kè 作文课	4				

Table 7.35 Use of 有 (yǒu) + *NOUN* in LIVT, TESOL and NTESOL

有(yǒu) + N	LLR	Sig.		
LIVT - TE	0.43	0.512		-
LIVT - NTE	3.76	0.053		-
TE - NTE	1.50	0.220		-

Table 7.36 Log-Likelihood Ratios: 有 (yǒu) + *NOUN* in LIVT, TESOL and NTESOL

When examining the nouns that accompany 有 (yǒu) with high frequency, the word bāngzhù 帮助 [help] turns out to be frequently used by all the three groups of participants. The words xìngqù 兴趣 [interest] and bànfǎ 办法 [method/means] were also often used by the LIVT and NTESOL participants. The frequent use of these nouns is related to what had been discussed during the interviews. We discussed whether the traditional drill method of learning Chinese classics and the essay class at school could do any good for students. We also discussed the participants' hobbies. As we mentioned previously, the pattern 有 (yǒu) followed by a noun/noun phrase is commonplace, expressing 'procession and existence. Therefore, it is not surprising to see this pattern and these nouns occurring in high frequencies in all these three groups. In our investigation on the colligation 有 (yǒu) with R1 nouns, we did not find great differences between these three different groups of Mandarin speakers.

有(yǒu) + Verbs

When examining the use of 有 (yǒu) with R1 colligate verbs, it is found that the TESOL participants used the pattern 有 (yǒu) + *VERB* more often than the LIVT and NTESOL participants. Repetitions of the word 有 (yǒu) were counted as being instances of this pattern and then we excluded these instances of stammering afterwards. After removing the repetitions (有(yǒu) + 有 (yǒu)), we found that the TESOL participants still showed greater use of 有(yǒu) + *VERB* in comparison with the LIVT and NTESOL participants. There are 86 occurrences of 有 (yǒu) + *VERB* in the LIVT data (2.23 times per 1000 words), 119 occurrences of 有 (yǒu) + *VERB* in the TESOL data (3.37 times per 1000 words), and 99 occurrences of 有 (yǒu) + *VERB* in the NTESOL data (2.35 times per 1000 words.) (See Table 7.37.) The TESOL participants used the pattern 有 (yǒu) + *VERB* more frequently than the LIVT and NTESOL participants, and this difference shows a statistical significance in log-likelihood ratios. (See Table 7.38.)

	LIVT			**TESOL**			**NTESOL**		
	Freq.	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**	**Freq.**	**%**	**Per 1000 words**
有 (yǒu)	870	100	22.56	870	100	24.64	1059	100	25.14
有 (yǒu) + V	109	12.53	2.83	120	13.79	3.40	112	10.58	2.66
有 (yǒu) +V (excluding 有 (yǒu))	86	9.89	2.23	119	13.68	3.37	99	9.35	2.35
Verbs with 有 (yǒu)	有 23 去 **9** 看 **8** 参加 **8** 说 6 写 6 看过 5 教 4			有 11 参加 **7** 去 **7** 教 6 写 6 说 5 背 5 让 4 讲到 4 听 4			有 13 看 **12** 去 **9** 参加 **7** 写 5 想 4		

Table 7.37 Use of 有 (yǒu)+ *VERB* in LIVT, TESOL and NTESOL

有 (yǒu) + V (excluding repeated 有 (yǒu))	LLR	Sig.		
LIVT - TE	**8.64**	**0.003**	**	-
LIVT - NTE	0.13	0.721		-
TE - NTE	**7.05**	**0.008**	**	+

Table 7.38 Log-Likelihood Ratios: 有 (yǒu)+ *VERB* in LIVT, TESOL and NTESOL

When we examining the verbs that accompany with 有 (yǒu) with high frequency, it is found the words 去 (meaning 'go') and 参加 [participate/join] to be frequently used by all the groups of participants. The word 看 [see/view] was also often used by the LIVT and NTESOL participants. In the interview, we discussed what kinds of events the participants went to. We also discussed the kinds of books/TV shows that these participants preferred to read/watch for leisure. (In Mandarin Chinese, the verbs 读 ('to read') and 看 ('to see/to view') can also be used when the object is 书 ('book').) It was therefore unsurprising to find these verbs frequently used by all the groups of participants. However, as we mentioned previously, the pattern 有 (yǒu) followed by a verb phrase is used mostly in Taiwan for expressing 'completed action/past' (and also 'experience', which will be discussed in the following section.) This use originated in the southern Chinese dialects (闽(mǐn)、粤(yuè)) and then influenced the use of 有 (yǒu) in Mandarin Chinese (Shi 1996, Zhang 1998, Chen 2007). The use of the 有 (yǒu) + VP pattern, previously used mainly in Taiwan (闽 (mǐn) dialect) and Hong Kong (粤 (yuè) dialect), has now spread to mainland China. (Sun, 2003) Some research even suggests that the use of English may play a part in the spreading use of the 有(yǒu)+VP pattern (Cui, 2013).

The three groups of participants, who were born and lived in Taiwan most of their time, were expected to apply the use of 有 (yǒu) + VP pattern in their speech. However, it is interesting to find that the TESOL group showed a significant difference in terms of the frequency of this pattern when compared with that of the other two groups of participants. It is worth investigating whether there is any difference in the manner of expressing 'completed action/past' and 'experience' amongst groups in their Mandarin Chinese. We will take the verb 参加 ('participate/join') for a closer look, because it is the most frequent verb shared by all the three groups in connection with the 有 (yǒu) + VP pattern. But in order to do this, there is a need to introduce another two words in Mandarin Chinese.

In Mandarin Chinese, there are two other words used to express tense and aspect (past/ perfect). One is the word 了, as we briefly mentioned previously. The other is the word 过. The former word has various functions. One of its major functions is as a completed action marker. (See Example (G).)

(G)	我 去 **了** 泰国 三次。
	Wǒ qù **le** tàiguó sāncì。
	I visited/went to Thailand three times.
(G')	我 **有** 去 泰 国 三 次。
	Wǒ **yǒu** qù tàiguó sāncì。
	I visited/went to Thailand three times.

The latter word 过 is an experience marker suggesting the thing/event happened in the past. (See Example (H).) Both of these markers are used after the verbs. The word 有(yǒu) can also be used as a completed action marker (Sample (G')) and an experience marker (Sample (H')). The word 有 (yǒu) is however used before the verbs. When the word 有(yǒu) is used as an experience marker, it can co-occur with the word 过. It rarely co-occurs with the word 了 when it functions as a completed action marker.

(H)	我 去 **过** 泰 国 。
	Wǒ qù **guò** tàiguó。
	I visited/was in Thailand.
(H')	我 **有** 去 泰 国 。 = 我 **有(yǒu)** 去 **过** 泰 国 。
	Wǒ **yǒu** qù tàiguó。 = Wǒ **yǒu** qù **guò** tàiguó。
	I visited/was in Thailand.

We checked the use of these two words 了[completed action marker] and 过[experience marker] with the verb 参加[participate]. The results show that all the three groups of participants did not use the word 了[completed action marker] with the verb 参加[participate] for expressing the completed action of 参加[participate]. Considering the questions which elicited the use of 参加 [participate] in the interview, we might expect to see 参加 [participate] used to talk about past experiences. (The participants were asked about their previous experiences of participating in language-relevant competition.) There are 18 instances of the use of 参加 with 'experience' markers 有 and 过 in the LIVT data. 8 out of 18 instances (44%) take the form of 有 (yǒu) 参加. 10 out of 18 (56%) take the form of 参加过. There are 11 instances of the use of 参加 with 'experience' markers in the TESOL data. 7 out of 11 instances (64%) take the form of 有(yǒu) 参加. 4 out of 11 (36%) take the form of 参加过. There are 14 instances of the use of 参加 with 'experience' markers in the NTESOL data. 7 out of 14 instances (50%) take the form of 有 (yǒu) 参加. 7 out of 14 (50%) take the form of 参加过. (See Table 7.39.)

	LIVT			TESOL			NTESOL		
	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words	Freq.	%	Per 1000 words
参加 (cānjiā)[participate]	23	--	0.60	27	--	0.76	40	--	0.95
Experience	18	100	0.47	11	100	0.31	14	100	0.33
有 (yǒu)* *参加 (cānjiā)	8	**44**	**0.21**	7	**64**	**0.20**	7	**50**	**0.17**
参加 (cānjiā) ***过 (guò)***	10	**56**	0.26	4	**36**	0.11	7	**50**	0.17
有 (yǒu) 参加 (cānjiā) 过 (guò)	*5*	--	*0.13*	*2*	--	*0.06*	*3*	--	*0.07*
Completed Action									
参加 (cānjiā) 了 (le)	0	0	0.00	0	0.	0.00	0	0	0.00

Table 7.39 Use of *参加(cānjiā)* in LIVT, TESOL and NTESOL

We can see that although the three groups of participants used the verb 参加[participate] with 'experience' markers in similar frequency[18], which markers they preferred were quite different. The TESOL participants tended to use the word 有 (yǒu) for expressing 'experience' rather than using the word 过. Such a tendency was not found in the LIVT and NTESOL participants. The use of 过 and 有 (yǒu) to express 'experience' in the LIVT and NTESOL groups were evenly split. The finding that the TESOL participants used the word 有 (yǒu) to express 'experience' instead of the word 过 and more often provides us with an explanation why the TESOL participants used the 有 (yǒu) + VP pattern significantly more often than the other groups of participants. The finding that the TESOL participants used the word 有(yǒu) for one particular discourse purpose more strongly than the other two groups of participants challenges our null hypotheses that there should be no noticeable differences in these participants' Mandarin use.

In the previous chapter, we found that the TESOL participants tended to use the English *is that* and the linking 是(shì) in 是你 [BE you] to connect noun phrases/clauses more often than the other participants. Taking account of the fact that the LIVT and TESOL participants were more likely to be exposed to English than the NTESOL participants, we thought there was a possibility that Mandarin speakers with such an exposure to English might be affected in their Mandarin Chinese by their primings of English use. Here we have another finding that partly supports this assumption. The special tendency to use 有 (yǒu) in the TESOL group may suggest that their extra English input (in

[18] The LIVT group had a higher frequency of the experience 参加 (0.47 times per 1000 words) in comparison with the TESOL and NTESOL groups (0.31/0.33 times per 1000 words). However, the difference is not statistically significant.

which they will be exposed to the auxiliary use of *have* in English) might have influenced their Mandarin Chinese use. However, if this English input had influence on how the TESOL participants used Mandarin Chinese, the fact that the LIVT participants did not use 有 (yǒu) + VP pattern with the same special tendency challenges the assumption. If, though, we take the language environment into account, it may be explainable.

The differences between the TESOL participants and LIVT participants are that the LIVT participants had been exposed to English input and had been producing English in an English-speaking environment. The TESOL participants had been exposed to English input and had been producing English in a Mandarin-speaking environment. The TESOL participants had also been exposed to abundant Mandarin Chinese input and been producing Mandarin Chinese in the same environment. If the way a person uses language is a result of constant adjustments based on the external comprehensible language input the person is exposed to and the meaningful output s/he produces as we had discussed in chapter 2, it may explain our findings. That is, the TESOL participants were receiving abundant Mandarin Chinese input and considerable English input, and they also were producing both Mandarin Chinese and English at the same time. The adjustment mechanism for language acquisition was, therefore, working for both Mandarin Chinese and English over the same period. If they received an English expression in a context which was similar to a context in which a Mandarin Chinese expression frequently occured and was used, it would be likely for the acquirers to fuse the priming of langauge use for these two expressions due to their similarity. Therefore, the result might lead the TESOL participants to be more likely to have a cross-language influence on their primings of language use.

The LIVT participants, on the other hand, had been receiving abundant comprehensible English input and they had had to produce an great amount of English output in an English-speaking environment at the same time. (Of course the LIVT participants continued using Mandarin in their daily lives, since there was a big Chinese community in the university, but the amount of Mandarin use was still incomparable to the amount of use in a Chinese-speaking environment.) The adjustment mechanism for language acquisition in this situation was likely working primary for English adjustment. The fusion might also occur to the LIVT participants, but only when the adjustment to the use of both English and Mandarin Chinese happened at the same time. Although this is for now an assumption and a hypothesis for explaining the special language use that we found in our investigation, it is an area worth studying and further research is required for the examination of this assumption. Nevertheless, what we have found in the use of 有 (yǒu) + VP pattern suggests that these were some factors that caused the TESOL group to behave differently from the other groups of

participants. The finding of noticeable differences in language use confirms that our fourth hypothesis of this study is incorrect.

7.3.3 Summary of the use of 有 (yǒu)

With regard to the use of the word 有 (yǒu), we found that some collocations were used significantly more often or less often by one or more groups of participants. The follow-up investigation showed that differences mostly lay in the quantity of use rather than in the differences of lexico-grammatical use. However, in the case of the collocations 还有 [still/in addition have/exist] and 还是有[still BE have], we found that the findings showed that our hypothesis 3 is incorrect. The findings are presented in Table 7.40.

Findings that Disagree with Hypothesis 3 (null hypothesis)			
Noticeable Differences	Collocations	The types of differences	Descriptions OR noteworthy findings/observation
The LIVT participants used the collocations **significantly less frequently** than the TESOL and NTESOL participants.	还有 (hái yǒu) [still/in addition have/exist]	Discourse Preference	The LIVT participants were not prone to use the collocation 还有 (hái yǒu) [still/in addition have/exist] in its connective function as often as the TESOL and NTESOL participants.
The LIVT participants used the collocations **significantly more frequently** than the TESOL and NTESOL participants.	还是有 (hái shì yǒu) [still BE have]	Discourse Preference	The LIVT participants used the collocation 还是 (hái shì) 有 (yǒu) [still BE have] more often than the TESOL and NTESOL participants for offering an idea that contradicts another speaker's idea

Table 7.40 Findings in the use of 有 (yǒu) [have/exist] that disagree with Hypothesis 3 (null hypothesis)

With regard to null hypothesis 4, the findings arising out of the investigation of the pattern 有 (yǒu) + *VERB* challenge this hypothesis. The TESOL participants used the pattern 有 (yǒu) + *VERB* more frequently than the LIVT and NTESOL participants to a statistically significant degree. (See Table 7.41.) The explanation of this special tendency cannot be a difference of focus or a difference in discourse behaviours. We can, however, see the possibility that the use of the 有 (yǒu) + *VERB* pattern in the TESOL data may be caused by the TESOL participants' extra English input in their TESOL study.

Findings that Disagree with Hypothesis 4 (null hypothesis)			
Noticeable Differences	Collocations	The types of differences	Descriptions OR noteworthy findings/observation
The TESOL participants used the collocations **significantly more frequently** than the LIVT participants.	还有 (hái yǒu) [still/in addition have/exist]	Discourse Preference	The TESOL participants were prone to use the collocation 还有(hái yǒu)[still/in addition have/exist] in its connective function more often than the LIVT participants.
The TESOL participants used the collocations **significantly more frequently** than the LIVT and NTESOL participants.	有 (yǒu) + VERB	Lexico-grammatical difference	The TESOL participants used the pattern 有 (yǒu) + VERB more often than the LIVT and TESOL participants. In their use of experience markers, the TESOL participants used the word 有 (yǒu) more often than the word 过 (guò). Such a strong tendency is not found in the LIVT and NTESOL groups.

Table 7.41 Findings in the use of 有(yǒu)[have/exist] that Disagree with Hypothesis 4 (null hypothesis)

7.4 Conclusion

In this chapter, we have examined two words: the English *have* (and all the *HAVE-verb* inflections) and the Chinese word 有 (yǒu) [have/exist]. For the first part of our investigation, we wanted to see whether the evidence supported our first and second hypotheses, namely that the LIVT participants, who were studying for at least one year in the UK, would have noticeable differences in their English use in comparison with the TESOL and NTESOL participants, who were studying in Taiwan, and also that the TESOL participants, who were studying English-relevant subjects in Taiwan, would have noticeable differences in their English use in comparison with those with non-English relevant subjects. The investigation on the use of *have* across the groups showed that several noticeable differences in terms of frequency existed. Therefore, the findings support our hypotheses 1 and 2. The findings are presented in Table 7.42 and Table 7.43.

Summary of the findings that Agree with Hypothesis 1 – Discourse Preference	
Collocations	Descriptions OR noteworthy findings/observation
I have *I have (to)*	The LIVT participants did not use the obligation *I have (to)* as often as the NTESOL participants did in their use of *I have*.
have to	The LIVT participants did not use the word *have* in its use of expressing obligation as often as the TESOL and NTESOL participants did in general.
Summary of the findings that Agree with Hypothesis 1 – Lexico-grammatical Difference	
Collocations	Descriptions OR noteworthy findings/observation
I've	The LIVT participants used *have* in its auxiliary verb function *(I + Aux. HAVE-verbs*) in the context of *I have('ve)* more often than the NTESOL participants. The LIVT participants abbreviated the word *have* when they used the word *have* for its auxiliary function in connection with *I* more often than the TESOL and NTESOL participants.

Table 7.42 Summary of the findings of the use of *have* that support Hypothesis 1

Summary of the findings that Agree with Hypothesis 2 – Discourse Preference	
Collocations	Descriptions OR noteworthy findings/observation
I have/ *I have (to)*	The TESOL participants did not use the obligation *have (to)* as often as the NTESOL participants did in their use of *I have*.
Summary of the findings that Agree with Hypothesis 2 – Lexico-grammatical Difference	
Collocations	Descriptions OR noteworthy findings/observation
I've	The TESOL participants used *have* in its auxiliary verb function *(I + Aux. HAVE-verbs)* in the context of *I have('ve)* more often than the NTESOL participants.

Table 7.43 Summary of the findings of the use of *have* that support Hypothesis 2

In our examination of the use of 有 (yǒu), we tested whether all the three groups showed no noticeable differences in their Mandarin Chinese use, given that they all had received similar

Mandarin Chinese input in Taiwan most of their lives. Their different English input backgrounds were predicted to have little influence on their Mandarin use. We found however that there were some noticeable differences in terms of the frequencies of certain words/word sequences. Therefore, our null hypotheses regarding their Mandarin use were incorrect. The findings are presented in Table 7.44 and Table 7.45.

Summary of the findings that disagree with Hypothesis 3 (null hypothesis) – Discourse Preference	
Collocations	Descriptions OR noteworthy findings/observation
还有 (hái yǒu) [still/in addition have/exist]	The LIVT participants did not use the collocation 还有 (hái yǒu) [still/in addition have/exist] in its connective function as often as the TESOL and NTESOL participants.
还是有 (hái shì yǒu) [still BE have]	The LIVT participants used the collocation 还是 (hái shì) 有 (yǒu) [still BE have] more often than the TESOL and NTESOL participants for offering an idea that contradicts another speaker's idea.
Summary of the findings that disagree with Hypothesis 3 (null hypothesis) – Lexico-grammatical Difference	
Collocations	Descriptions OR noteworthy findings/observation
	N/A

Table 7.44 Summary of the findings of the use of 有 (yǒu) [have/exist] that disagree with Hypothesis 3

Summary of the findings that disagree with Hypothesis 4 (null hypothesis) – Discourse Preference	
Collocations	Descriptions OR noteworthy findings/observation
还有 (hái yǒu) [still/in addition have/exist]	The TESOL participants used the collocation 还有 (hái yǒu) [still/in addition have/exist] in its connective function more often than the LIVT participants.
Summary of the findings that disagree with Hypothesis 4 (null hypothesis) – Lexico-grammatical Difference	
Collocations	Descriptions OR noteworthy findings/observation
有 (yǒu) + VERB	The TESOL participants used the pattern *有 (yǒu) + VERB* more often than the LIVT and TESOL participants. In connection with the use of experience markers, the TESOL participants used the word 有 (yǒu) more often than the word 过 (guò). Such a strong tendency is not found in the LIVT and NTESOL groups.

Table 7.45 Summary of the findings of the use of 有 (yǒu) [have/exist] that disagree with Hypothesis 4

We found two noticeable differences which suggest that there may be some influences from the English input. The LIVT participants tended to abbreviate the word *have* when it was accompanied by pronouns. The application of the abbreviated *have ('ve)* was found much more frequently in the LIVT group than in the TESOL and NTESOL groups. The use of *'ve* is a common use in the reference corpora (authentic spoken English.) Therefore, this finding suggests that the LIVT participants was

possibly influenced by the English input to which they had exposed. The finding of the investigation on the pattern *有 (yǒu) + VERB* also raised the possibility that the TESOL participants were influenced by the English input. They used the *有 (yǒu)+VP* pattern in their Mandarin more frequently than the other groups of participants. These findings, to some degree, are potentially compatible with what had been discussed in chapter 2. The differences in English input to which the participants were exposed may cause some differences in their language use.

Chapter 8 Conclusions

8.1 Conclusions of the current study

In this study, we have focused on the issue of whether Taiwanese EFL learners with different English learning backgrounds (environment/education) have noticeable different language use in both spoken English and spoken Mandarin. We formulated our research hypotheses based on the idea that language input is one of the key elements to second language acquisition. Krashen's Input hypothesis (1982, 1985, 1998) is the starting point of this study. We took the view which deems people's acquisition of language starts from collecting/mimicking lexical items/multi-word units which are obtained from communicative/comprehensible input, and people use their inborn cognitive ability, pattern-forming ability, to generate rules of patterns/frames/formulaic sequences from frequently-used language instances. In this study, we intended to investigate whether the EFL learners with the different kinds of English learning backgrounds (immersion in the authentic English environment and studying in English-relevant subjects in non-English-speaking environment) will have significant differences in language use. In our discussion, we tried to see whether the findings are compatible with what had been discussed in chapter 2 (the importance of language input/output) and whether the lexical priming theory can be used for possible explanation. The concept of lexical priming theory (Hoey, 2005, 2008) is that when people acquire language, they acquire habits of usage derived their experience of frequently hearing what others do in particular contexts. Based on the contextualised information, their minds are primed to associate each word/word sequence with the use of particular collocations, colligations, semantic associations. People may be primed to use particular words together or avoid such a usage in certain contexts. This language behaviour is based on their recognition that some words or combinations of words have co-occurred more frequently or less frequently in the course of their previous experience. With different kinds of input in different contexts, people will be primed to use certain expression more strongly or less strongly.

The first two hypotheses in this study concerned whether different English learning environments and different educational trainings in English would influence how the EFL learners used their spoken English. The third and fourth hypotheses concerned whether native Mandarin speakers whose Mandarin input was similar but who had different English learning backgrounds would show little difference in their spoken Mandarin. These four hypotheses are as follows:

H1. Because of the new language input, which they will receive in an authentic English environment, those EFL learners who go to the UK for study will show

noticeable differences in their English use, in comparison with those who stay in Taiwan

H2. Due to the diversity of their school education and undergraduate and postgraduate studies, the EFL learners studying in English-relevant subjects will show noticeable differences in their use of English from those studying non-English-relevant subjects.

H3. Due to learning Chinese as their first language in a similar natural Chinese-speaking environment, those EFL learners who travel to the UK for study will exhibit no noticeable differences in their Mandarin Chinese use when compared with those who stayed in Taiwan.

H4. For the same reason, those EFL learners studying in English-relevant subjects will show no noticeable differences in their use of Mandarin Chinese than those studying in non-English-relevant subjects.

In order to test our hypotheses, three types of Taiwanese EFL learners were involved in this study. The first group, which I have referred to throughout the thesis as the LIVT group, had lived and studied in the UK (Liverpool) for at least one year, and they were likely to have received authentic and abundant English input in their daily lives and in their studies, in contrast with the other participants. The second group, which I have referred to as the TESOL group, were Taiwanese postgraduate students studying English-relevant subjects (mostly TESOL) in Taiwan. They were living in a Mandarin-speaking country where English is a foreign language, so their opportunities for hearing and using English in their daily lives were relatively few. They needed to use English in the classroom. Therefore, the TESOL participants had comparatively more opportunities to use English than ordinary learners of English in Taiwan. The third group, which I have referred to as the NTESOL group (the control group), were Taiwanese postgraduate students who had received the college level of formal English education at school. They were also living in a Mandarin-speaking country with little chance to use English in their daily lives. They were not majoring in English-relevant subjects. Although they might still use some English for their studies, their English input was very limited. I interviewed these participants individually, and asked them the same set of questions (both in English and in Mandarin.) Their use of spoken English and spoken Mandarin was our materials for testing our four hypotheses.

Our investigations concerned the use of the English pronouns *I*, *you* and *it* (in chapter 4) and the Mandarin pronouns 我[wǒ][I], 你[nǐ][you] and 它[tā][it] (in chapter 5.) We also investigated the use of two sets of English and Mandarin verbs, namely *is* and 是[shì][BE] (in chapter 6) and *have* and 有[yǒu][have/exist] (in

chapter 7.) The analyses were conducted to see which L1 collocates and R1 collocates of these words were most commonly used by these participants. When the use of these collocations (containing the L1 and R1 collocates) were found to be significantly more frequent or less frequent in the speech of one or more of the groups, we then undertook a further investigation into how these collocations were used by the participants.

There were two kinds of differences that we investigated in our analyses. We tried to identify whether the noticeable difference was in the domain of discourse differences or was in the domain of lexico-grammatical differences. The latter kind of differences concerned differences in the use of a collocation, colligation or semantic association of words/word sequences. When one particular use of words/word sequences was significantly more or less frequent in one particular group, we considered this a noticeable difference, and interpreted whether the potential causes for this difference were relevant to the speaker's language input. Our main aim for this study was to answer whether EFL learners with different English learning environments/training backgrounds behaved differently in their language use (rather than why). We did not attempt to explore the exact causes. There were other potential influences (e.g. social/educational factors) that would need to be considered if we wanted to discover what exactly caused these differences in the discourse/lexico-grammatical features of language use. What we tried to achieve was point out the differences in language use which would potentially provide us with grounds for further research.

What we have found in our investigations showed that our hypotheses 1 and 2 were supported. There were some noticeable differences in the use of English between groups. In our investigation into the use of English pronouns *I, you* and *it*, we discovered that the LIVT participants used some words/word sequences for discourse purposes more often than the other two groups of participants. For instance, the LIVT participants used *I just (just)* more often for adding extra emphasis or hedging in their speech. They also used *I think* more frequently as a discourse strategy to earn more time to generate the appropriate words for their sentences. We also discovered some English use in the LIVT group which suggested the possibility of input influences. For instance, the use *like* before *if you* in the LIVT group showed a great similarity to the use found in the reference corpora (BASEah and COCAsp). In spoken English, the word *like* in that context often functions similarly as a connective for linking ideas or as a filler that helps the speaker find time for their next words. The LIVT participants were prone to use *like* in the same way as the native speakers of the reference corpora, and the TESOL and NTESOL participants did not show such a tendency. These noticeable differences, therefore, indicated our hypothesis 1 is correct.

With respect to hypothesis 2, we also found that the TESOL participants avoided using the pronoun *it* when they used the verb *use*, while the NTESOL participants used the collocation *use it* often in their spoken English. The TESOL participants spelled out the object in full when using the verb *use* more often than did the other participants. This special tendency might have been influenced by their English training (in academic and written style), which required a relatively more precise description in language use.

In connection with the English verbs *is* and *have*, we also found some evidence which supported hypotheses 1 and 2. We found that the LIVT participants used the adverb *quite* (*is('s) quite)* to modify their adjectives more frequently than the other participants, and likewise, the TESOL participants used the adverb *pretty* (*is('s) pretty*) to modify their adjectives more frequently than the other participants. These two findings appeared to be the effect of different preferences of words. However, it was not clear why different preferences should be associated with different linguistic experience. In addition, we also found a lexico-grammatical difference in the TESOL participants' English use. The TESOL participants used *is that* in the pattern *NP+ is that + [S]* more frequently than the LIVT and NTESOL participants. With regard to the use of *have*, we found that both the LIVT and TESOL participants used *have* in its auxiliary verb function (*I + Aux. HAVE-verbs*) in the context of *I have('ve)* more often than the NTESOL participants. In addition, the LIVT participants abbreviated the word *have* when they used the word *have* for its auxiliary function with *I* more often than the TESOL and NTESOL participants. This tendency in the LIVT group might have been influenced by their English input in which the abbreviated *have* (*'ve*) had been commonly used. All of these findings again supported our hypotheses 1 and 2.

Our investigation of hypotheses 1 and 2 showed that there were significant differences between the EFL learners with different English learning/training backgrounds in their use of spoken English. However, when we tested our hypotheses 3 and 4 (null hypotheses), we also found that there were significant differences between the EFL learners with different English learning/training backgrounds in their use of spoken Mandarin. Our hypotheses 3 and 4, therefore, were proved to be incorrect. In our investigation on the use of the Mandarin pronouns 我 [I], 你 [you] and 它 [it] (nearly identical in meaning and function to the English pronouns), we found that the LIVT and TESOL participants had several kinds of use that differed from those of the other participants. The LIVT participants, similarly to what they did with English pronouns, showed a strong tendency to use certain words as a discourse means for adding extra emphasis, downtoning or hedging or as a strategy for earning more time to generate their utterances. These kinds of discourse behaviours were also used by the other participants. However, the tendency was not as strong as that of the LIVT participants'. For instance,

in the use of 你 说 [you say], the LIVT participants used the pattern '*verb + 你 说 [you say]+ NP/[S]*' in which the word *说 (shuō) [to say]* was used as an emphasiser relatively more often than the other participants. Some of our findings also implied there might be a possibility of influence from their different English language input. For instance, the LIVT participants consistently used the words *I just/just* and 我 就 [I just]/ 就 [just] as a discourse strategy in both English and Mandarin. This consistent use of a similar English and Mandarin words for this purpose was comparatively rarely seen in the language use of the other participants, especially in that of the NTESOL participants. This finding suggests that for the LIVT participants, who had been exposed to comparatively more abundant and comprehensible English input and who had had more opportunities to practice meaningful output, the division between their primings of certain English words and of roughly equivalent Mandarin words which share some similarity in use, had become less distinct. That is to say, there might be a cross-language influence in people's primings of language use, not in one way (English input influences the use of English), but in two ways (English input influences the use of English and Mandarin). A similar situation was also found with regard to the TESOL participants' language use. The TESOL participants used 它 [it] as an object after a verb (了解 它 [to understand it]) less frequently than the other participants. This finding matched what we discovered in the use of *use it* and *learn it* in 4.4.2, in which the TESOL participants also showed a low use of *it* as an object. This consistency in the use of English *it* and Mandarin 它 [it] suggests that their primings of these two words could have been influenced across languages. We found more evidence suggesting that cross-language primings might exist when we investigated the use of 是 [to be] and 有 [to have/exist]. In our investigation on the use of 是 你 [BE you], we found that the TESOL participants used the pattern '*noun phrase + 是 (shì) + noun phrase / clause*' more frequently than the other groups of participants. This greater use of the pattern by the TESOL participants matched our finding with regard to their use of *is that*. In these two patterns, the word 是 (shì) and the words *is (that)* were used for linking noun phrases/clauses. The fact that the TESOL participants showed greater use of the linking 是 (shì) and the linking word sequence *is that* than the other participants supported our speculation about the existence of cross-language primings of language use. Our finding in the investigation on the pattern *有 (yǒu) + VERB* also raised the possibility that the TESOL participants had been influenced by their English input. The TESOL participants used the '*有 (yǒu) + VP'* pattern in their Mandarin more frequently than the other groups of participants. This special tendency with regard to the use of 有 (yǒu) [have/exist] suggested that the TESOL participants' extra English input (in which they were exposed to the auxiliary use of *have* in English) might have influenced their Mandarin Chinese. All of these findings lead us to assume the existence

of cross-language primings of language use. However, the current study could only point to the possibility. Much more research on a wider range of items and a larger body of data would be needed to provide convincing proof.

8.2 Limitations of this research and Suggestions for the future research

To discover whether different input can influence how people use language, a strictly input-controlled subject is essential. This research was designed to collect data from participants whose English learning backgrounds meant they could be grouped as appropriate subjects. Nevertheless, there remained a concern that the language input the participants within each category had received was not sufficiently similar. In fact, we could not know definitively what kind of English/Mandarin Chinese input they had received. Thus, although we categorised the participants according to an expectation of differences associated with majors (TESOL and non-TESOL) and study locations (in Taiwan and in the UK), it was highly likely that other factors informed the participants' use of language. Extra curricular activities, such as watching TV, travelling or surfing online, could also be sources of input. Besides, psychological factors, such as motivation, are also crucial in terms of learning a language. Therefore, a survey regarding these factors would be required, which unfortunately was not done in this study. Despite the fact that we could still use the corpus described to test the hypotheses, it would been more valid to posit additional instances for comparison and generalisation. Since it was the relationship between learners' English input (their contact with English) and their spoken output (the choice of word use) that we aimed to explore in this study, a more controlled and more comparable EFL learner corpus regarding their language contact would be required for the further research.

For the purpose of finding out whether different language input can influence the way people are primed to use language, we need input-controlled participants. In this research we tried to recruit many participants with English learning experience similar to the experience of their fellow group members, we were still concerned that the language input which these participants received was not similar enough. Therefore, in future research it would be desirable to make the input difference more distinct and precise and to have a long period of observation of the subjects. A better understanding of the English input to which the EFL learners have been exposed, such as the content of their course books or English media they encounter in their daily lives, would be required. It would also be ideal to observe a group of EFL learners on the same course and then to observe whether and how their language use changes after having been given different types of input during their courses. An observation of two groups of EFL learners undertaking different types of English input would also be

a good way to see how the different input (educational training) might affect people's language use. It would be ideal to test how the language use of EFL learners in a different kind of native speaker environment (similar to the LIVT group in our study) would be affected. Since the examination would require fluent users of both English and Mandarin, advanced EFL learners would still be the ideal choice. Therefore, it would be good to approach EFL learners who are planning to go on working holidays in the UK or Australia. The duration of a working holiday is usually one year. It would be possible to have three interviews for each subject: one before they go to the UK/Australia, one in the middle of their working holiday and one after they are back to motherland. In this way, it would be possible to observe whether the authentic English-speaking environment might affect their use of English and how. It would also be possible to observe whether their use of native language becomes different after they return. Future research should consider having a fourth interview after the participants have been back to a native language-speaking environment for more than 6 months. If the evidence for cross-language effects on native language disappears after 6 months, it would both strengthen the case for saying that the language environment is more likely to be the factor that influences both Mandarin and English and also indicate that there are no lasting negative effects of the influence.

What we had found in the current study suggested that there might be a possibility in which the newly-learnt English primings might influence a Mandarin speaker's original Mandarin primings. What is now needed is to test whether cross-language influences on primings would be found in the opposite direction. It would therefore be valuable to have a similar study of how CFL learners (Learning Chinese as a Foreign Language) use spoken Mandarin and spoken English. Three groups of participants would also be needed: a group of CFL learners (native English speakers) who have been in a Mandarin Chinese-speaking environment for at least one year, a group of CFL learners who have majored in Chinese-relevant subjects (in an English-speaking environment,) and a group of CFL learners who are not in a Mandarin-Chinese speaking environment and are not studying or working in a Mandarin Chinese-rich area. If noticeable differences are found in the English of the three groups of CFL learners, then the existence of cross-language effects on primings would be further supported. If the cross-language effects on primings exist, we should also be able to find the evidence not merely in studies of English and Mandarin, but in studies of all kinds of languages. More research on speakers using different types of language (e.g. one inflectional (semi-inflectional) language (Spanish, German) versus a non-inflectional language (Mandarin, Japanese) or a comparison between two inflectional/non-inflectional languages) would be needed.

With regard to our assumption of cross-language influence on primings, when we discussed the use of 还有 (hái yǒu) [still/in addition have/exist] in 7.3.1, we identified several possibilities of English

influence on the use of the connective 还有 [in addition/and also] in the LIVT group. One of the possibilities was that the contexts in which the connective 还有 [in addition/and also] is characteristically used might be contexts in which the English connectives are rarely used. Therefore, the LIVT participants might have been influenced to avoid using the connective (还有) in such contexts. To investigate this, it would be necessary to identify what kind of contexts are associated with the Mandarin use and to compare these with the kind of contexts associated with the English use, and then see how (and whether) the subjects use the two languages in the expected contexts. That is to say, a context-based analysis would be required.

Another possibility was that their primings of English connectives might influence the subjects to use other kinds of Mandarin connectives more frequently rather than the connective 还有 [in addition/and also] to organise their texts/ideas. In this case, it would be necessary to observe how the subjects use the connectives in general and how they organise their ideas with these connectives. That is to say, a text-based comparison between the use of English and Mandarin would be required. It would also be valuable if we could collect representative samples of the subjects' written English and written Mandarin to see whether there are noticeable differences in how they use written English/Mandarin, and whether the potential cross-language effects on primings found in the spoken languages could also be found in the written languages.

What we have just suggested shows that there are ways of improving the research offered in this research (both in the materials and the methods) and further avenues of investigation for future researchers.

Appendices

Appendix 1.
Top 30 frequent wordlists

	LIVT		TESOL		NTESOL	
N	**Word**	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**
1	**I**	**1826**	**I**	**1915**	EH	1932
2	EH	1485	EH	1905	**I**	**1657**
3	THE	1179	THE	1216	THE	910
4	TO	805	TO	985	TO	772
5	**YOU**	**714**	AND	787	AND	599
6	YEAH	666	IN	507	ENGLISH	441
7	THINK	599	ENGLISH	489	IN	434
8	AND	564	SO	480	YEAH	430
9	IN	559	THINK	478	THINK	406
10	SO	451	A	477	**IT'S**	**375**
11	**IT'S**	**432**	**IS**	**407**	SO	359
12	ENGLISH	414	THEY	390	**IS**	**358**
13	THEY	414	**IT'S**	**382**	BECAUSE	313
14	**IS**	**360**	YEAH	370	THEY	312
15	BUT	346	OF	331	A	310

The top 30 English words in the LIVT, TESOL and NTESOL groups (1)

	LIVT		TESOL		NTESOL	
N	**Word**	**Freq.**	**Word**	**Freq.**	**Word**	**Freq.**
16	A	342	BECAUSE	329	BUT	281
17	LIKE	332	BUT	316	YES	260
18	JUST	321	**YOU**	**299**	**HAVE**	**250**
19	BECAUSE	311	OR	292	**YOU**	**236**
20	THAT	278	THAT	291	OR	218
21	**IT**	**267**	FOR	289	LIKE	213
22	FOR	249	LIKE	276	CAN	207
23	OR	245	**HAVE**	**264**	**IT**	**202**
24	OF	232	NOT	254	JUST	194
25	DON'T	230	**IT**	**248**	NOT	186
26	**HAVE**	**227**	CAN	238	MY	181
27	CAN	222	MY	230	OF	177
28	NOT	221	YES	221	SCHOOL	171
29	MY	218	SCHOOL	220	WE	169
30	IF	213	WHEN	212	VERY	165

The top 30 English words in the LIVT, TESOL and NTESOL groups (2)

	LIVT		**TESOL**		**NTESOL**	
N	Word	Freq.	Word	Freq.	Word	Freq.
1	的 de	1841	的 de	1718	的 de	2130
2	**我 wǒ**	**1711**	**我 wǒ**	**1270**	**我 wǒ**	**1585**
3	**是 shì**	**1127**	**是 shì**	**941**	就是 jiùshì	1214
4	就是 jiùshì	1046	就是 jiùshì	916	**是 shì**	**1139**
5	**你 nǐ**	**1014**	**有 yǒu**	**870**	**有 yǒu**	**1059**
6	那 nà	901	就 jiù	829	对 duì	893
7	**有 yǒu**	**870**	对 duì	802	不 bú/bù	805
8	对 duì	819	那 nà	752	会 huì	797
9	不 bú/bù	817	会 huì	660	觉得 juédé	781
10	就 jiù	796	不 bú/bù	642	就 jiù	746
11	个 gè/ge	749	**你 nǐ**	**632**	那 nà	736
12	会 huì	705	觉得 juédé	559	**你 nǐ**	**637**
13	觉得 juédé	685	个 gè/ge	549	个 gè/ge	620
14	啊 ā	646	然後 ránhòu	539	然後 ránhòu	560
15	然後 ránhòu	476	啊 ā	468	嗯 èn	518

The top 30 Mandarin Chinese words in the LIVT, TESOL and NTESOL groups (1)

	LIVT		**TESOL**		**NTESOL**	
N	Word	Freq.	Word	Freq.	Word	Freq.
16	可能 kěnéng	451	很 hěn	398	**它 tā**	**515**
17	因为 yīnwèi	402	要 yào	375	因为 yīnwèi	497
18	一 yī/y	397	因为 yīnwèi	375	很 hěn	437
19	什么 shénme	370	说 shuō	343	一 yī/yì	428
20	很 hěn	363	一 yī/yì	328	啊 ā	423
21	沒 méi	344	什么 shénme	323	沒 méi	410
22	这 zhè	344	看 kàn	321	看 kàn	401
23	写 xiě	338	**它 tā**	**296**	要 yào	348
24	说 shuō	332	比较 bǐjiào	293	也 yě	342
25	**它 tā**	**331**	也 yě	282	在 zài	320
26	看 kàn	327	嗯 èn	282	什么 shénme	319
27	要 yào	298	写 xiě	250	所以 suóyǐ	304
28	也 yě	274	沒 méi	245	说 shuō	300
29	在 zài	272	去 qù	244	比较 bǐjiào	293
30	种 zhǒng	268	都 dōu	239	可能 kěnéng	287

The top 30 Mandarin Chinese words in the LIVT, TESOL and NTESOL groups (2)

Appendix 2.

Sample transcriptions and transcription conventions for corpus data

[LIVT01] English transcriptions (Answers to first three questions)

yeah yes I do I like English &eh because I think English is a a kind of language tool for for me to communicate with foreigners and also &eh sometimes when we read &eh kind of other book written in English word the drama or the films are shoot in English and I mean the language they use &eh even sometimes &eh they translate &eh from translate their language into our L_one &chi &eh Mandarin_Chinese but .
I would say &eh there's no word equivalent into you knowl between L_one L_two so &eh I think if we we understand English we can just go to go back to their original version because in this way we can understand &eh the meaning of it and behind it because you know sometimes we just read the translator and we can't get the meaning yeah so yeah I do I do like speaking English yeah .
yeah exactly .
okay so &eh when I was like junior high school students I I have no chance because it's a compulsory subject at first and but &eh but now okay so I when I was in Taiwan because it's it's &eh especially &eh in when I was in junior high school and senior high school it was a a subject &com in our compulsory education so I had to learn of course but &eh .
I chose English as my major when I was in university because &eh after I mean when I was in senior high school I found &eh I fall in love in English actually so that's why I chose my major in university I don't know why probably because I kind of &eh attracted by you know , western culture probably one of the reason right and &haha those things sound silly &eh I thought like if you could speak English it sound it sounds cool &haha so yeah .

right , western culture .
why &eh maybe okay so maybe I would talk about a little bit answer a little bit &eh the answer of in question three .
it won't won't get problem right okay so maybe I can answer three together is that okay ?
so &eh I will answer like the English I learn in junior high school and senior high school bef(ore) at first &eh when I was in junior high school &eh because as you know &eh , the education system in Taiwan is quite examination oriented and so everything I mean English other I mean our knowledge of English I learn in junior high school was only focus like &eh grammatical accuracy and vocabulary as well and our teacher that time they they didn't &eh tell us we need to focus on &eh listening skills and speaking as well &eh but I was so I I think I was really lucky when I was in senior high school because I went to quite a private school and we our school had a kind of program like &eh you gonna pay &eh pay for extra money for &eh like school kind of was &eh I call it it's after after_school section &eh like we gonna stay in in another &eh extra two hours after &s the normal class and that that kind of programs &eh is train kind of train you to &eh to be proficient in English area so &eh we I used to remember &eh I have like two two conversation class a week and two grammars a week so I would say &eh at that time I start to have a a lot of opportunity to get language input so at that time I also started to &eh I don't know get &impo I mean be exposed to &eh American culture as well like watch watch a lot of American films or listen to American songs I I guess maybe start from that time I kind of you know get interested in western culture so that's the reason I probably I think uhhuh .

[LIVT01] Transcription conventions (English) (Answers to first three questions)

#_FO yeah_UH yes_UH I_PPIS1 do_VD0 I_PPIS1 like_VVI English_JJ &eh;
because_CS I_PPIS1 think_VV0 English_NN1 is_VBZ a_AT1 a_AT1 kind_NN1 of_IO language_NN1 tool_NN1 for_CS for_IF me_PPIO1 to_TO communicate_VVI with_IW foreigners_NN2 and_CC also_RR &eh; sometimes_RT when_CS we_PPIS2 read_VV0 &eh; kind_NN1 of_IO other_JJ book_NN1 written_VVN in_II English_JJ word_NN1 the_AT drama_NN1 or_CC the_AT films_NN2 are_VBR shoot_VV0 in_II English_NN1 and_CC I_PPIS1 mean_VV0 the_AT language_NN1 they_PPHS2 use_VV0 &eh; even_RR sometimes_RT &eh; they_PPHS2 translate_VV0 &eh; from_II translate_VV0 their_APPGE language_NN1 into_II our_APPGE Lone_JJ χ &eh; MandarinChinese_NP2 but_CCB ._.
#_FO I_ZZ1 would_VM say_VVI &eh; there_EX 's_VBZ no_AT word_NN1 equivalent_NN1 into_II you_PPY knowl_VV0 between_II Lone_JJ Ltwo_NN1 so_CS &eh; I_PPIS1 think_VV0 if_CS we_PPIS2 we_PPIS2 understand_VV0 English_NN1 we_PPIS2 can_VM just_RR go_VVI to_TO go_VVI back_RP to_II their_APPGE original_JJ version_NN1 because_CS in_II this_DD1 way_NN1 we_PPIS2 can_VM understand_VVI &eh; the_AT meaning_NN1 of_IO it_PPH1 and_CC behind_II it_PPH1 because_CS you_PPY know_VV0 sometimes_RT we_PPIS2 just_RR read_VV0 the_AT translator_NN1 and_CC we_PPIS2 ca_VM n't_XX get_VVI the_AT meaning_NN1 yeah_UH so_RR yeah_UH I_PPIS1 do_VD0 I_PPIS1 do_VDI like_II speaking_VVG English_NN1 yeah_UH ._.
#_FO yeah_UH exactly_RR ._.
#_FO okay_RR so_RR &eh; when_CS I_PPIS1 was_VBDZ like_II junior_JJ high_JJ school_NN1 students_NN2 I_ZZ1 I_PPIS1 have_VH0 no_AT chance_NN1 because_CS it_PPH1 's_VBZ a_AT1 compulsory_JJ subject_NN1 at_RR21 first_RR22 and_CC but_CCB &eh; but_CCB now_RT okay_RR so_CS I_PPIS1 when_CS I_PPIS1 was_VBDZ in_II Taiwan_NP1 because_CS it_PPH1 's_VBZ it_PPH1 's_VBZ &eh; especially_RR &eh; in_II when_RRQ I_PPIS1 was_VBDZ in_II junior_JJ high_JJ school_NN1 and_CC senior_JJ high_JJ school_NN1 it_PPH1 was_VBDZ a_AT1 a_AT1 subject_NN1 &com; in_II our_APPGE compulsory_JJ education_NN1 so_CS I_PPIS1 had_VHD to_TO learn_VVI of_RR21 course_RR22 but_CCB &eh; ._.
#_FO I_ZZ1 chose_VVD English_JJ as_CSA my_APPGE major_JJ when_CS I_PPIS1 was_VBDZ in_II university_NN1 because_CS &eh; after_CS I_PPIS1 mean_VV0 when_RRQ I_PPIS1 was_VBDZ in_II senior_JJ high_JJ school_NN1

I_PPIS1 found_VVD &eh; I_PPIS1 fall_VV0 in_II love_NN1 in_II English_NN1 actually_RR so_RR that_DD1 's_VBZ why_RRQ I_PPIS1 chose_VVD my_APPGE major_JJ in_II university_NN1 I_PPIS1 do_VD0 n't_XX know_VVI why_RRQ probably_RR because_CS I_PPIS1 kind_RR21 of_RR22 &eh; attracted_VVN by_II you_PPY know_VV0 ,_, western_JJ culture_NN1 probably_RR one_MC1 of_IO the_AT reason_NN1 right_NN1 and_CC &haha; those_DD2 things_NN2 sound_VV0 silly_JJ &eh; I_PPIS1 thought_VVD like_II if_CSW you_PPY could_VM speak_VVI English_NN1 it_PPH1 sound_VV0 it_PPH1 sounds_VVZ cool_RR &haha; so_RR yeah_UH ._.

#_FO right_RR ,_, western_JJ culture_NN1 ._.

#_FO why_RRQ &eh; maybe_RR okay_RR so_RG maybe_RR I_PPIS1 would_VM talk_VVI about_II a_AT1 little_JJ bit_NN1 answer_VV0 a_AT1 little_JJ bit_NN1 &eh; the_AT answer_NN1 of_IO in_II question_NN1 three_MC ._.

#_FO it_PPH1 wo_VM n't_XX wo_VM n't_XX get_VVI problem_NN1 right_RR okay_RR so_RG maybe_RR I_PPIS1 can_VM answer_VVI three_MC together_RL is_VBZ that_DD1 okay_JJ ?_? #_FO so_CS &eh; I_PPIS1 will_VM answer_VVI like_II the_AT English_NN1 I_PPIS1 learn_VV0 in_II junior_JJ high_JJ school_NN1 and_CC senior_JJ high_JJ school_NN1 bef(ore)_RT at_II first_MD &eh; when_CS I_PPIS1 was_VBDZ in_II junior_JJ high_JJ school_NN1 &eh; because_CS as_CSA you_PPY know_VV0 &eh; ,_, the_AT education_NN1 system_NN1 in_II Taiwan_NP1 is_VBZ quite_RG examination_NN1 oriented_VVD and_CC so_RR everything_PN1 I_PPIS1 mean_VV0 English_JJ other_JJ I_PPIS1 mean_VV0 our_APPGE knowledge_NN1 of_IO English_NN1 I_PPIS1 learn_VV0 in_II junior_JJ high_JJ school_NN1 was_VBDZ only_RR focus_VV0 like_II &eh; grammatical_JJ accuracy_NN1 and_CC vocabulary_NN1 as_RR21 well_RR22 and_CC our_APPGE teacher_NN1 that_DD1 time_NNT1 they_PPHS2 they_PPHS2 did_VDD n't_XX &eh; tell_VVI us_PPIO2 we_PPIS2 need_VV0 to_TO focus_VVI on_II &eh; listening_JJ skills_NN2 and_CC speaking_VVG as_RR21 well_RR22 &eh; but_CCB I_PPIS1 was_VBDZ so_RG I_MC1 I_PPIS1 think_VV0 I_PPIS1 was_VBDZ really_RR lucky_JJ when_CS I_PPIS1 was_VBDZ in_II senior_JJ high_JJ school_NN1 because_CS I_PPIS1 went_VVD to_II quite_RG a_AT1 private_JJ school_NN1 and_CC we_PPIS2 our_APPGE school_NN1 had_VHD a_AT1 kind_NN1 of_IO program_NN1 like_II &eh; you_PPY gon_VVGK na_TO pay_VVI &eh; pay_VV0 for_IF extra_JJ money_NN1 for_IF &eh; like_JJ school_NN1 kind_RR21 of_RR22 was_VBDZ &eh; I_PPIS1 call_VV0 it_PPH1 it_PPH1 's_VBZ after_II afterschool_NN1 section_NN1 &eh; like_CS we_PPIS2 gon_VVGK na_TO stay_VVI in_RP in_II another_DD1 &eh; extra_JJ two_MC hours_NNT2 after_II &s; the_AT normal_JJ

class_NN1 and_CC that_CST that_DD1 kind_NN1 of_IO programs_NN2 &eh; is_VBZ train_NN1 kind_RR21 of_RR22 train_VV0 you_PPY to_II &eh; to_TO be_VBI proficient_JJ in_II English_JJ area_NN1 so_CS &eh; we_PPIS2 I_PPIS1 used_VMK to_TO remember_VVI &eh; I_PPIS1 have_VH0 like_II two_MC two_MC conversation_NN1 class_NN1 a_AT1 week_NNT1 and_CC two_MC grammars_NN2 a_AT1 week_NNT1 so_CS I_PPIS1 would_VM say_VVI &eh; at_II that_DD1 time_NNT1 I_PPIS1 start_VV0 to_TO have_VHI a_AT1 a_AT1 lot_NN1 of_IO opportunity_NN1 to_TO get_VVI language_NN1 input_NN1 so_RR at_II that_DD1 time_NNT1 I_PPIS1 also_RR started_VVD to_II &eh; I_PPIS1 do_VD0 n't_XX know_VVI get_VV0 &impo; I_PPIS1 mean_VV0 be_VBI exposed_VVN to_II &eh; American_JJ culture_NN1 as_RR21 well_RR22 like_II watch_NN1 watch_VV0 a_AT1 lot_NN1 of_IO American_JJ films_NN2 or_CC listen_VV0 to_II American_JJ songs_NN2 I_ZZ1 I_PPIS1 guess_VV0 maybe_RR start_VV0 from_II that_DD1 time_NNT1 I_PPIS1 kind_RR21 of_RR22 you_PPY know_VV0 get_VV0 interested_JJ in_II western_JJ culture_NN1 so_RR that_DD1 's_VBZ the_AT reason_NN1 I_PPIS1 probably_RR I_PPIS1 think_VV0 uhhuh_NN1 ._.

[LIVT01] Mandarin transcriptions (Answers to first three questions)

嗯 我 不 太 喜歡，因為 我 覺得 不 實用，這 這 應該 是 我 認為 最 大 的 原因，因為 我，基本 上 我 一 開始 認為 不 實用 我 就 覺得，我 會 蠻 排斥 去 學習 它 ・

而且 我 認為，它 在 我們 求學 的 過程 好像 就 變成 只是 為了 升學 或 考試 的 目的 這 都 會 讓 我 更 更加 排斥，對，所以 就 我 個人 來 說 是 覺得 不 喜歡 ・

呃，我 個人 是 比較 不 喜歡 國文 或 歷史 這 方面 的 ・

就是 (...) 就 就是，一 開始 是 喜歡 不 喜歡，當 你 沒 有 興趣 你 就 會 更，更 排斥 接觸 這 方面 的 資訊，對 ・

對對對，可是，喔，可是 如果 相對 文言文 比如說 一些 什麼，現代 的，文章，新詩，散文 或者 是 那 樣子 我 是 還可以 接受 就是 文言文 我 是 不 會 接受 的，因為 我 覺得 它 是，不 實用 不 能 傳達 你 意見 的 ・

對對對，我 我 我 我 蠻 喜歡 讀 新詩 的，因為 我 覺得 它 可以，感受 作者 寫 詩 當下 的 心情 或 他 想 傳達 的 意思，即便 你，吸 接收 的 跟 他 當初 想 傳達 的 不 一樣 但是 它 還是 可以 可以 是 還是 可以 啟發 你 的 思想，對 ・

我 覺得 這 個 幫助 是 因為 我們，對 中國 歷史 來 說 可能，因為 中國 歷史 博大精深 嘛，可能 有 些 人 會 覺得 說 古 時候 有 傳 著 很多 道理，那 這些 道理 或 這些 知識 能 藉由 古文 被 分享，那 要是 我們 能 記 著 這些 資訊，常常 複習 比如說 古詩 或者 是 古文，那，或許 我們 能 從 這些 過程 之中 熟知 那些 大 道理，在 特殊 的 情況 下 我們 會 能 領悟 出 那些 道理，所以 我 覺得 或多或少 有 幫助 吧？

我 覺得 這 個，就是 學習 國文 或者 是，一 個 歷史 來 說 我 覺得 這 是 必須 的，就 即便 你 當下，對 小朋友 來 說 當下 他 可能 只是 興趣 覺得 背 那 個 是 有趣，或者 是 被 強迫，但是 等 之後 或許 你 有方 到 用 時 一 天 就 像，就 像 我們 第一 題 說 你 會 不 會 用到，我 覺得 或許 那 是 或多或少 的 幫助 了 ・

[LIVT01] Transcription conventions (Mandarin) (Answers to first three questions)

#(FW) 嗯(I) 我(Nh) 不(D) 太(Dfa) 喜歡(VK) ,(COMMACATEGORY) 因為(Cbb) 我(Nh) 覺得(VK) 不(D) 實用(VH) ,(COMMACATEGORY) 這(Nep) 這(Nep) 應該(D) 是(SHI) 我(Nh) 認為(VE) 最(Dfa) 大(VH) 的(DE) 原因(Na) ,(COMMACATEGORY) 因為(Cbb) 我(Nh) ,(COMMACATEGORY) 基本(Na) 上(Ng) 我(Nh) 一(D) 開始(VL) 認為(VE) 不(D) 實用(VH) 我(Nh) 就(D) 覺得(VK) ,(COMMACATEGORY) 我(Nh) 會(D) 蠻(Dfa) 排斥(VC) 去(D) 學習(VC) 它(Nh) ・(PERIODCATEGORY)

#(FW) 而且(Cbb) 我(Nh) 認為(VE) ,(COMMACATEGORY) 它(Nh) 在(P) 我們(Nh) 求學(VA) 的(DE) 過程(Na) 好像(D) 就(D) 變成(VG) 只是(D) 為了(P) 升學(VA) 或(Caa) 考試(Na) 的(DE) 目的(Na) 這(Nep) 都(D) 會(D) 讓(VL) 我(Nh) 更(D) 更加(Dfa) 排斥(VC) ,(COMMACATEGORY) 對(VH) ,(COMMACATEGORY) 所以(Cbb) 就(D) 我(Nh) 個人(Na) 來(D) 說(VE) 是(SHI) 覺得(VK) 不(D) 喜歡(VK) ・(PERIODCATEGORY)

#(FW) 呃(D) ,(COMMACATEGORY) 我(Nh) 個人(Na) 是(SHI) 比較(Dfa) 不(D) 喜歡(VK) 國文(Na) 或(Caa) 歷史(Na) 這(Nep) 方面(Na) 的(DE) ・(PERIODCATEGORY)

#(FW) 就是(Cbb) ((PARENTHESISCATEGORY) ・(PERIODCATEGORY) ・(PERIODCATEGORY) ・(PERIODCATEGORY))(PARENTHESISCATEGORY) 就(D) 就是(Cbb) ,(COMMACATEGORY) 一(D) 開始(VL) 是(SHI) 喜歡(VK) 不(D) 喜歡(VK) ,(COMMACATEGORY) 當(P) 你(Nh) 沒有(VJ) 興趣(Na) 你(Nh) 就(D) 會(D) 更(D) ,(COMMACATEGORY) 更(D) 排斥(VC) 接觸(VC) 這(Nep) 方面(Na) 的(DE) 資訊(Na) ,(COMMACATEGORY) 對(VH) ・(PERIODCATEGORY)

#(FW) 對對對(VH) ,(COMMACATEGORY) 可是(Cbb) ,(COMMACATEGORY) 喔(I) ,(COMMACATEGORY) 可是(Cbb) 如果(Cbb) 相對(VH) 文言文(Na) 比如說(P) 一些(Neqa) 什麼(Nep) ,(COMMACATEGORY) 現代(VH) 的(DE) ,(COMMACATEGORY) 文章(Na) ,(COMMACATEGORY) 新詩(Na) ,(COMMACATEGORY) 散文(Na) 或者(Caa) 是(SHI) 那(Nep) 樣子(Na) 我(Nh) 是(SHI) 還可以(D) 接受(VC) 就是(Cbb) 文言文(Na) 我(Nh) 是(SHI) 不會(D) 接受(VC) 的(DE) ,(COMMACATEGORY) 因為(Cbb) 我(Nh) 覺得(VK) 它(Nh) 是(SHI) ,(COMMACATEGORY) 不(D) 實用(VH) 不能(D) 傳達(VD) 你(Nh) 意見(Na) 的(DE) ・(PERIODCATEGORY)

#(FW) 對對對(VH) ,(COMMACATEGORY) 我(Nh) 我(Nh) 我(Nh) 我(Nh) 蠻(Dfa) 喜歡(VK) 讀(VC) 新詩(Na) 的(DE) ,(COMMACATEGORY) 因為(Cbb)

我(Nh) 覺得(VK) 它(Nh) 可以(D) ,(COMMACATEGORY) 感受(Na) 作者(Na) 寫(VC) 詩(Na) 當下(D) 的(DE) 心情(Na) 或(Caa) 他(Nh) 想(VE) 傳達(VD) 的(DE) 意思(Na) ,(COMMACATEGORY) 即便(Cbb) 你 (Nh) ,(COMMACATEGORY) 吸(VC) 接收(VC) 的(DE) 跟(P) 他(Nh) 當初 (Nd) 想(VE) 傳達(VD) 的(DE) 不(D) 一樣(VH) 但是(Cbb) 它(Nh) 還是(D) 可以(D) 可以(D) 是(SHI) 還是(D) 可以(D) 啟發(VC) 你(Nh) 的(DE) 思想 (Na) ,(COMMACATEGORY) 對(VH) ・(PERIODCATEGORY)

#(FW) 我(Nh) 覺得(VK) 這(Nep) 個(Nf) 幫助(Na) 是(SHI) 因為(Cbb) 我們 (Nh) ,(COMMACATEGORY) 對(P) 中國(Nc) 歷史(Na) 來(D) 說(VE) 可能 (D) ,(COMMACATEGORY) 因為(Cbb) 中國(Nc) 歷史(Na) 博大精深(VH) 嘛 (T) ,(COMMACATEGORY) 可能(D) 有(V_2) 些(Nf) 人(Na) 會(D) 覺得(VK) 說古(VA) 時候(Na) 有(V_2) 傳(VD) 著(Di) 很多(Neqa) 道理 (Na) ,(COMMACATEGORY) 那(Dk) 這些(Neqa) 道理(Na) 或(Caa) 這些(Neqa) 知識(Na) 能(D) 藉由(P) 古文(Na) 被(P) 分享(VJ) ,(COMMACATEGORY) 那 (Dk) 要是(Cbb) 我們(Nh) 能(D) 記(VC) 著(Di) 這些(Neqa) 資訊 (Na) ,(COMMACATEGORY) 常常(D) 複習(VC) 比如說(P) 古詩(Na) 或者 (Caa) 是(SHI) 古文(Na) ,(COMMACATEGORY) 那 (Dk) ,(COMMACATEGORY) 或許(D) 我們(Nh) 能(D) 從(P) 這些(Neqa) 過 程(Na) 之中(Ng) 熟知(VK) 那些(Neqa) 大(VH) 道理 (Na) ,(COMMACATEGORY) 在(P) 特殊(VH) 的(DE) 情況(Na) 下(Ng) 我們 (Nh) 會(D) 能(D) 領悟出(VJ) 那些(Neqa) 道理(Na) ,(COMMACATEGORY) 所以(Cbb) 我(Nh) 覺得(VK) 或多或少(D) 有(V_2) 幫助(VC) 吧 (T) ?(QUESTIONCATEGORY)

#(FW) 我(Nh) 覺得(VK) 這(Nep) 個(Nf) ,(COMMACATEGORY) 就是(Cbb) 學 習(VC) 國文(Na) 或者(Caa) 是(SHI) ,(COMMACATEGORY) 一(Neu) 個(Nf) 歷史(Na) 來(D) 說(VE) 我(Nh) 覺得(VK) 這(Nep) 是(SHI) 必須(D) 的 (DE) ,(COMMACATEGORY) 就(D) 即便(Cbb) 你(Nh) 當下 (D) ,(COMMACATEGORY) 對(P) 小朋友(Na) 來(D) 說(VE) 當下(D) 他(Nh) 可能(D) 只是(D) 興趣(Na) 覺得(VK) 背(VC) 那(Nep) 個(Nf) 是(SHI) 有趣 (VH) ,(COMMACATEGORY) 或者(Caa) 是(SHI) 被(P) 強迫 (VF) ,(COMMACATEGORY) 但是(Cbb) 等(VC) 之後(Ng) 或許(D) 你(Nh) 有 方(VH) 到(P) 用(VC) 時(Ng) 一(Neu) 天(Nf) 就(D) 像 (VG) ,(COMMACATEGORY) 就(D) 像(P) 我們(Nh) 第一(Neu) 題(Nf) 說(VE) 你(Nh) 會不會(D) 用到(VC) ,(COMMACATEGORY) 我(Nh) 覺得(VK) 或許 (D) 那(Nep) 是(SHI) 或多或少(D) 的(DE) 幫助(VC) 了(Di) ・ (PERIODCATEGORY)

Bibliography

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological review*, *102*(1), 68.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, *2*, 89-195.

Bialystok, E., & Fröhlich, M. (1978). Variables of classroom achievement in second language learning. *The Modern Language Journal*, *62*(7), 327-336.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language 68(4)*, 706-747.

Bruner, J. (1990). *Acts of Meaning*. Harvard University Press.

Bruner, J. (2009). Culture, mind and education. *Contemporary Theories of Learning*. Routledge, 159-68.

Brook, G. L. (1960). *A history of the English language 2nd edition*. A. Deutsch.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18(1),* 32-42.

Bryson, B. (1991). *Neither here nor there: travels in Europe*. Secker & Warburg.

Butler, C. (1985). *Statistics in linguistics*. Blackwell.

Butler, C. (2004). Corpus studies and functional linguistic theories. *Functions of Language*, *11(2)*, 147-86.

Carter, R. (1988). Vocabulary, cloze and discourse: An applied linguistic view. *Vocabulary and language teaching*, 161-180.

Chen, Y. (2007). Research on the source of "YOU+VP" structure from the south dialect's formation. *Journal of Gansu Lianhe University (Social Science Edition)*, (04).

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory*, *2*(3), 113-124.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chen, K-J. and Huang, C-R. (2015). “Academia Sinica Balanced Corpus”, in: *Encyclopedia of Chinese Language and Linguistics*, General Editor Rint Sybesma. Consulted online on 10 August 2018. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000191

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240-247.

Collins, A. M., & Quillian, M. R. (1970). Does category size affect categorization time?. *Journal of verbal learning and verbal behavior*, *9*(4), 432-438.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. *Cognition in Learning and Memory*. Wiley.

Cook, G. (2003). *Applied Linguistics.* Oxford: Oxford University Press.

Cressie, N. and Read, T. R. C. (1984). Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society*. *Series B (Methodological)*, 46(3), 440-464.

Cui, N. (2013). "有+VP" pattern in modern mandarin. *Journal of Yunnan Normal University Teaching and Research on Chinese as a Foreign Language, (04)*.

Day, R. (1985). The use of the target language in context and second language proficiency. *Input in second language acquisition*, 257-271. Newbury House.

Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

de Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. *Language and Computers*, *33*, 51-68.

de Mornay Davies, P. (1998). Automatic semantic priming: The contribution of lexical- and semantic-level processes. *European Journal of Cognitive Psychology 10(4),* 389-412.

DeKeyser, R. M. (1986). *From Learning to Acquisition?: Foreign Language Development in a US Classroom and During a Semester Abroad*. University Microfilms.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. *Focus on form in classroom second language acquisition*, 42-63.

DeKeyser, R. M. (2001). Automaticity and automatization. *Cognition and second language instruction,* 125-151. Cambridge University Press.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, *46*(4), 613-642.

Duffy, T. M. & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. *Handbook of Research for Educational Communications and Technology*, Macmillan Library Reference USA.

Dulay, H., Burt, M. & Krashen, S. (1982). *Language Two*. Oxford University Press.

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics,* 19(1) March, 61-74.

Ellis, N. C. (1996). Sequencing in SLA. *Studies in second language acquisition*, *18*(1), 91-126.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in second language acquisition*, *24*(2), 143-188.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The handbook of second language acquisition*, 63-103.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *Tesol Quarterly*, 42(3), 375-396.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1). *Studies in Corpus Linguistics (SCL)*, 89.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in second language acquisition*, *21*(2), 285-301.

Felix, S. W., & Weigl, W. (1991). Universal grammar in the classroom: The effects of formal instruction on second language acquisition. *Interlanguage studies bulletin (Utrecht)*, *7*(2), 162-180.

Ferguson, C. A. (1995). Foreword to the first edition. *Second language acquisition in a study abroad context (Vol. 9)*, xi. John Benjamins Publishing.

Fillmore, C. J. (1977). The case for case reopened. *Syntax and Semantics 8: Grammatical Relations*, 59-82. Academic Press.

Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. *Frames, fields, and contrasts: New essays in semantic and lexical organization*, *103*, 75-102.

Freed, B. F. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. *Georgetown University round table on languages and linguistics*, 459-477.

Freed, B. F. (Ed.). (1995). *Second Language Acquisition in a Study Abroad Context* (Vol. 9). John Benjamins Publishing.

Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis* (Vol. 241). University of California Press.

Gries, S. T. (2005). Syntactic priming. A corpus-based approach. *Journal of Psycholinguistic Research* 34(4), 365-399.

Hammond, R. M. (1988). Accuracy versus communicative compentency: The acquisition of grammar in the second language classroom. *Hispania*, *71*(2), 408-417.

Han, Z. (2004). *Fossilization in adult second language acquisition* (Vol. 5). Multilingual Matters.

Hoey, M. (2004a). Lexical priming and the properties of text. *Corpora and Discourse, 385-412.*

Hoey, M. (2004b). The textual priming of lexis. *Corpora and language learners*, *17*, 21.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge.

Hoey, M. (2008). Lexical priming and literary creativity. *Text, Discourse and Corpora: Theory and Analysis*, 7-29. Continuum.

Hopper, P. (1988). Emergent grammar and the a priori grammar postulate. *Linguistics in context: Connecting observation and understanding*, 117-134.

Hopper, P. (1998). Emergent grammar. *New Psychology of Language*, 155-175.

Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.

Irujo, S. (1986). A piece of cake:/earning and teaching idioms. *ELT journal*, *40*(3), 236-242.

Jung, J. Y. (2002). Issues in acquisitional pragmatics. *Teachers College-Columbia University Working Papers in TESOL and Applied Linguistics*, *2*(2), 1-34.

Kachru, B. B. (1982). South Asian English. *English as a world language*, *4*.

Kachru, B. B. (2006). The English language in the outer circle. *World Englishes*, *3*, 241-255.

Kellerman, E. (1978). Giving learners a break: Native language intuitions as a source of predictions about transferability. *Working Papers on Bilingualism Toronto*, (15), 59-92.

Kemmer, S., &. Barlow, M (2000). Usage-based models of language.

Kemmer, S., & Israel, M. (1994). Variation and the usage-based model. *CLS*, *30*, 165-179.

Krashen, S. D. (1976). Formal and informal linguistic environments in language acquisition and language learning. *Tesol Quarterly*, 157-168.

Krashen, S. D. (1982). *Child-Adult Differences in Second Language Acquisition. Series on Issues in Second Language Research*. Newbury House Publishers, Inc., Rowley, MA 01969.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.

Krashen, S. D. (1998). Comprehensible output?. *System*, *26*(2), 175-182.

Krashen, S. D. (2003). *Explorations in language acquisition and use*. Heinemann.

Krashen, S. D., & Seliger, H. W. (1976). The role of formal and informal environments in second language learning: A pilot study. *Linguistics*, *14*(172), 15-22.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom*. Pergamon Press.

Lado, R. (1964). *Language teaching: a scientific approach*. McGraw-Hill.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.

Langacker, R. W. (1988). A usage-based model. *Topics in Cognitive Linguistics*, 127-161. Benjamins.

Langacker, R. W. (2000). A dynamic usage-based model. *Usage-based models of language*, 1-63.

Langacker, R. W. (2002). *Concept, image, and symbol: The cognitive basis of grammar* (Vol. 1). Walter de Gruyter.

Lewis, M. (1993). The lexical approach: The state of ELT and a way forward. *Language Teaching Publications*.

Lightbown, P. M., Spada, N., Ranta, L., & Rand, J. (1999). *How languages are learned* (Vol. 2). Oxford University Press.

Lightbown, P. M., & Spada, N. (2013). *How Languages are Learned 4th edition-Oxford Handbooks for Language Teachers*. Oxford University Press.

Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York academy of sciences*, *379*(1), 259-278.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input1. *Applied linguistics*, *4*(2), 126-141.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*, *41368*.

Ma, W. Y., & Chen, K. J. (2003). Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. In *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17* (pp. 168-171). Association for Computational Linguistics.

MacWhinney, B. (Ed.). (1999). *The emergence of language*. Taylor & Francis.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk 3rd Edition*. Lawrence Erlbaum Associates.

MacWhinney, B. (2001). Emergentist approaches to language. *TYPOLOGICAL STUDIES IN LANGUAGE*, *45*, 449-470.

McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: an introduction*. Edinburgh University Press.

Meyer, D. E., & Schvanefeldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227-234.

Mitchell, R. & Myles, F. (2004). *Second language learning theories 2nd edition*. Oxford University Press.

Nattinger, J. R., & DeCarrico, J. S. (1989). Lexical phrases, speech acts and teaching conversation. *Aila Review*, *6*, 118-139.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Nation, P. (2001). *Learning Vocabulary In Another Langauge*. Cambridge: Cambridge University Press.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory processes. *Memory & Cognition* 4(5). 648-54.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1-135.

Nikolov, M., & Krashen, S. (1997). Need we sacrifice accuracy for fluency?. *System*, *25*(2), 197-201.

Nunan, D. (2012). *What is this thing called language?*. Palgrave Macmillan.

Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh University Press.

Pace-Sigge, M. (2013). *Lexical priming in spoken English usage*. Springer.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, *191*, 225.

Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. *7th International Conference on Statistical analysis of textual data (JADT 2004)* , 926-936.

Ross, C. & Ma, J. S. (2006) *Modern Mandarin Chinese Grammar: A Practical Guide*. Routledge.

Ross, C. & Ma, J. S. (2014) *Modern Mandarin Chinese Grammar: A Practical Guide* (2nd edition). Routledge.

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL quarterly*, 41-51.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research 6*, 93-124.

Schachter, J. (1996). Maturation and the issue of Universal Grammar in second language acquisition. *Handbook of second language acquisition*, 159-193.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use* (Vol. 9). John Benjamins Publishing.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. *Formulaic sequences: Acquisition, processing and use*, 1-22.

Schwartz, B. D. (1998). The second language instinct. *Lingua*, *106*(1-4), 133-160.

Scovel, T. (2001). *Learning new languages: A guide to second language acquisition*. Boston: Heinle & Heinle.

Scott, M. (2012). *WordSmith Tools version 6*. Lexical Analysis Software.

Seliger, H. W. (1977). Does practice make perfect?: A study of interaction patterns and L2 competence1. *Language Learning*, *27*(2), 263-278.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209-232.

Sherman, S.J., Crawford, M.T., Hamilton, D.L. and Garcia-Marques, L. (2003). Social inference and social memory: the interplay between systems. *The SAGE Handbook of Social Psychology*, 45-68. SAGE Publications Ltd.

Shi, C. (1996). 论“有”字句. *语言研究, (1),* 28-33.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, J. (1997). The lexical item. *Contrastive Lexical Semantics*, 1-24. Benjamins.

Sinclair, J. (2000). Lexical grammar. *Naujoji Metologija 24*, 191-203.

Sinclair, J. (2004). *Trust the Text. Language, Corpus and Discourse*. Routledge.

Skinner, B. F. (1957). The experimental analysis of behavior. *American scientist*, *45*(4), 343-371.

Spada, N. (1986). The interaction between type of contact and type of instruction: Some effects on the L2 proficiency of adult learners. *Studies in Second Language Acquisition*, *8*(2), 181-199.

Spada, N., & Lightbown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in second language acquisition*, *15*(2), 205-224.

Stern, H. H. (1983). *Fundamental Concepts of Language Teaching: Historical and Interdisciplinary Perspectives on Applied Linguistic Research*. Oxford University Press.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, *2*(1), 23-55.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell publishers.

Sun, Q. (2003). The "有+ VP" sentence pattern in dialogues. *Journal of Chinese Language and Culture School Nanjing Normal University, (03)*.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, *15*, 165-179.

Swain, M. (1995). Three functions of output in second language learning. *Principle and practice in applied linguistics: Studies in honour of HG Widdowson*, *2*(3), 125-144.

Swain, M. (2005). The output hypothesis: Theory and research. *Handbook of research in second language teaching and learning*, *1*, 471-483.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, *16*(3), 371-391.

Teubert, W. (2005). My version of corpus linguistics. *International journal of corpus linguistics*, *10*(1), 1-13.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

Thompson, P. and Nesi, H. (2001) The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research* 5 (3) 263-264

Towell, R., & Hawkins, R. D. (1994). *Approaches to second language acquisition*. Multilingual matters.

Vygotsky, L. S. (1962). Piaget's Theory of Child Language and Thought.

Vygotsky, L. S. (1978). Interaction between learning and development. *Readings on the development of children*, *23*(3), 34-41.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *SSLA*, 27, 33-52.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 Learners. *Studies in Second Langauge Acquisition*, 30(1), 79-95.

White, L. & A. Juffs. (1998). Constraints on wh-movement in two different contexts of non-native language acquisition: competence and processing. *The Generative Study of Second Language Acquisition*, 111-130. Lawrence Erlbaum.

Winitz, H. (1996). Grammaticality judgment as a function of explicit and implicit instruction in Spanish. *The Modern Language Journal*, *80*(1), 32-46.

Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal*, *20*(1), 01-15.

Wray, A. (1992). *The focusing hypothesis: the theory of left hemisphere lateralised language re-examined* (Vol. 3). John Benjamins Publishing.

Wray, A. (2002). *Formulaic sequences and the lexicon*. Cambridge University Press.

Xu, J. (2009). Log-likelihood ratio calculator. Beijing: National Research Center for Foreign Language Education. Beijing Foreign Studies University.

Zhang, Y. (1998). “有”字句研究综述. *汉语学习, (3),* 28-32.